



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

보건학석사학위논문

진균류 동정을 위한 마커 유전자 기반
검색시스템 구축에 관한 연구

Fungi Classification
Based on Genetic Markers

2013년 8월

서울대학교 보건대학원
보건학과 바이오인포매틱스전공
장 진 화

국문 초록

다양한 생물들이 공존하고 있는 생태환경에서 진균류는 숙주세포에 기생하며 생태계의 분해자 역할을 수행하고 인간 생활에 있어 식품의 원료가 되기도 하며 질환을 일으키는 병원균이기도 하다. 최근 곰팡이 균의 감염사례 증가와 함께 생활환경의 변화로 인한 피부병 발생빈도가 증가하고 있고 약해진 면역력으로 인해 재발경우가 많을 뿐만 아니라 완치가 어렵기에 진균증에 대한 연구의 필요성은 증대되고 있다. 매우 빠른 속도로 증식하고 다른 생명체에까지 전파력이 강한 진균 등장의 가능성은 향후 보건학적으로 큰 이슈가 될 것이다. 이에 따라 진균의 생활사 및 유전학적 연구는 보건학적으로도 필수적이다. 지구상 예측되는 진균류의 개체수는 150만 중 임에도 불구하고 현재까지 동정된 종은 10만 종에 그친다. 분자생물학의 발달은 새로운 진균류 유전자 서열 정보의 축적을 가져왔으며 이들의 계통분류학적 정보를 얻기 위해 현재까지 알려진 진균류의 유전자 서열 데이터들을 체계적으로 저장할 수 있는 데이터베이스의 구축은 필수적이다. 현재까지 전체 지놈 서열이 밝혀진 진균류는 1,000여 종에 불과하지만 NGS 방법으로 인해 앞으로 더욱 빠르게 유전체 서열을 얻을 수 있을 것이다. 하지만 아직까지 동정되지 않은 진균류들의 유전자 서열 정보와 이들이 가지고 있는 유전자들의 기능은 완벽하게 알려지지 않았다. 그리고 수많은 진균류를 분류할 수 있는 유전자 마커의 부재로 인해 새로운 진균류의 종 동정 시 분류 체계 정립에 어려움이 존재한다. 따라서 본 연구에서는 진균류의 동정을 위한 마커 유전자 기반 검색 시스템을 구축하고 이를 기반으로 계통분류학적 연구를 수행하였다. 구축한 데이터베이스는 진균류의 유전자 마커로 주로 이용된 상위 6개의 유전자 서열 정보 및 분류체계 정보를 통합하였다. 구축한 검색 시스템은 <http://lcbb.snu.ac.kr/fgdb/> 에서 확인 할 수 있으며 FGDB라고 명명하였다.

본 검색시스템을 이용하면 유전자 마커에 따라 진균류의 서열정보 및 분류체계 정보를 얻을 수 있고 독자적으로 구축한 마커유전자 기반의 데이터베이스에서 BLAST 웹 인터페이스를 구축하여 새로운 진균류의 분류체계를 알고 싶은 쿼리 서열 정보를 입력하면 상동성이 높은 상위 100개의 진균 종의 Accession 넘버와 분류체계 정보를 얻을 수 있다. 또한 본 데이터베이스를 기반으로 진균류의 분류가 효율적으로 이루어 지는 Multigene 마커의 조합을 찾아내는 연구를 수행하여 Atpase6, Rpb2, β -Tub의 유전자 조합이 가장 높은 신뢰도를 보이며 진균류를 대상으로 이들 마커 조합을 이용한 Classification이 잘 이루어졌음을 확인 하였다. 또한 본 연구에서는 바이오인포매틱스 기법을 바탕으로 진균류의 보건학적 연구에 적용하고자 하였다. 특히 구축한 검색 시스템을 이용하여 새롭게 등장하는 감염성 진균의 계통분류학적 위치를 확인하여 분류학적 위치에 따른 항진균제 치료 기법을 적용하면 인구 집단에선 진균증의 빠른 확산을 막을 수 있을 것이다. 뿐만 아니라 국내외 진균들의 계통분류학 체계를 보완하고 매년 밝혀지는 새로운 진균의 빠른 종 동정이 가능할 것이다. 본 연구를 기반으로 하여 계통분류체계 정립의 마련은 분류 군에 따라 질병을 일으키는 진균류의 항 진균제 타겟 연구에 기반이 되어 보건학 발전에 기여할 수 있을 것이라 기대한다.

주요어 : 진균, 진균증, 바이오인포매틱스, 데이터베이스, 계통분류, 유전자 마커, 보건.

학 번 : 2011-23863

목 차

국문 초록	i
목 차	iii
표 목차	v
그림 목차	vi

I 서 론

1. 연구 배경

1.1 진균의 종류 및 분류체계의 변화	1
1.2 진균류 유전정보 관련 데이터베이스	6
1.3 진균에 의한 감염성 질병	10
1.4 진균의 계통수	17

2. 연구의 필요성	22
------------------	----

3. 연구 목적	24
----------------	----

II. 연구 방법

1. 데이터 수집 및 가공	26
2. 시스템 개발 환경 및 BLAST 서버 구축	30

3. 데이터베이스 구현	32
4. 데이터베이스 분석 항목.....	36
5. 계통수 작성과 정확도 측정	39

III. 연구 결과

1. 마커 유전자 기반 검색 시스템 구축	41
2. 데이터베이스 기반 Multigene 마커 탐색	54

IV. 고 찰

1. 활용 방안	66
2. 보건학적 연구에 적용	68
3. 향후 보완 및 추가계획	70
4. 기대성과	72

V. 결 론

참고문헌	78
Abstract	83
감사의 글	85

표 목 차

Table 1.1 Database for fungal research	9
Table 1.2 Symptom of mycoses	16
Table 2.1 System development environment	31
Table 2.2 Amount of data constructed on BLAST database	31
Table 2.3 Schema of the database	34
Table 2.4 Analysis systems for making multigene	38
Table 3.1 Amount of sequence data	43
Table 3.2 Species of multigene dataset	56
Table 3.3 Accuracy of each phylogenetic tree	65

그림 목차

Figure 1.1 The total number of fungi present in the world	5
Figure 1.2 The total number of fungi genome sequence completed	5
Figure 1.3 Seven phylums based on AFTOL project	19
Figure 1.4 Classification of <i>S.cerevisiae</i>	21
Figure 2.1 Genbank data before processing	28
Figure 2.2 Genbank data after processing using Java scripts	29
Figure 2.3 Data flow diagram	35
Figure 3.1 Front page of the database	45
Figure 3.2 Search table of marker gene menu	46
Figure 3.3 Search table of multigene menu	48
Figure 3.4 Standalone BLAST web interface	50
Figure 3.5 Result of species identification using BLAST	51
Figure 3.6 Pair-wise alignment web interface	52
Figure 3.7 Result of pair-wise alignment	53

Figure 3.8 Result of insertion sequence using SequenceMatrix	57
Figure 3.9 Phylogenetic tree(NJ) based on Rpb gene	59
Figure 3.10 Phylogenetic tree(NJ) based on Efl- α gene	59
Figure 3.11 Phylogenetic tree(NJ) based on β -Tub gene	60
Figure 3.12 Phylogenetic tree(NJ) based on Atpase 6 gene.....	60
Figure 3.13 Phylogenetic tree(NJ) based on rRNA gene	61
Figure 3.14 Phylogenetic tree(NJ) based on Atpase, Rpb, β -Tub gene.....	61
Figure 3.15 Phylogenetic tree(NJ) based on rRNA, Atpase, Rpb, β -Tub gene	63
Figure 3.16 Phylogenetic tree(NJ) based on rRNA, Atpase, Rpb, β -Tub, Efl- α gene	63

I. 서 론

1. 연구 배경

1.1 진균류 특징 및 분류체계의 변화

진균류는 곰팡이, 효모, 버섯 등을 포함하며 2004년 전 세계적으로 7만2천 종이 보고되었고(Fryar *et al.*, 2004) 2008년에 9만7천 종이 알려졌으며 우리나라에는 3만7천 종 정도가 있을 것이라 추정하고 있다(Silvia *et al.*, 2008). 여기에 보고되지 않은 진균까지 포함하면 지구상에 약 150만종의 진균이 존재할 것이라 예측되고 있다(Hawksworth *et al.*, 1991). 이렇게 진균류의 높은 종 다양성은 생태계에 진균류가 생물학적 다양성과 생태계 순환과정에서 중요한 부분을 차지하고 있음을 말해 준다(Fryar *et al.*, 2004). <Figure 1.1>은 예측되는 진균류의 숫자에 비해 종 이름의 동정이 완료된 진균류가 전체의 약 6%에 불과하다는 것을 보여주고 있다. 분자 생물학이 발달하기 이전, 진균류는 주로 생활주기, 배우자, 배우자낭, 포자과, 포자 등의 구조적 특징 같은 세포학적 및 형태적 특징에 기초하여 분류되었다. 자낭균류(*Ascomycota*)는 긴 자루모양의 자낭속에 포자를 생성하며 담자균류(*Basidiomycota*)는 담자기 아래 담자포자가 있으며 유성생식을 한다. 병꼴균류(*Chytridiomycota*)는 물에서 생활하며 포자 끝에 한 개의 편모로 헤엄쳐 앞으로 나아가며 죽은 수생식물에 기생한다. 접합균류(*Zygomycota*)는 토양, 동물, 식물 사체에서 생활하며 무성생식, 유성생식 모두 가능하고 격벽이 없어 다핵 균사체가 발달되었다. 진균은 이러한 문(Phylum) 아래 세부적으로 나뉘어 지는 강(Class), 목(Order), 과(Family) 등도 각각의 형태학적 특징에 따라 나뉘게 된다(Elizabeth *et al.*, 2005). 진균류의 유전자 서열 정보는 분자생물학적 기술이 발달함에 따라 축적되고 있으며 현재까

지 Genome 분석이 완료된 진균류의 개수는 1996년 처음으로 *S.cerevisiae*가 밝혀진 이래로 꾸준히 증가하고 있으며 2012년까지 1000개의 Genome 서열 분석을 완료하고자 1000 Fungal Genome Project가 계획되었다. <Figure 1.2>는 1996년부터 최근까지 년도에 따라 Genome 서열 분석이 완료된 진균 종의 개수로 향후에는 NGS(Next Generation Sequencing)의 발달로 진균류의 Genome 분석이 더욱 빠르게 이루어 질 것이다.

진균류의 분류에 기본이 되는 형태학적 분류체계는 수십만 종에 이르는 진균류 계통분류로의 한계를 갖는다. 그 이유는 수많은 종을 체계적으로 분류하는데 있어 형태학적 특징이 종의 수에 따라 다양하게 관찰되지 않기 때문이다. 또한 진균 종을 배양하는 과정에서 유전형이 바뀌거나 배양 시간이 오래 걸리는 단점을 가진다. 따라서 현재에는 형태학적 분류와 함께 유전자 서열을 기반으로 한 분류체계가 주를 이루고 있다. 분자생물학적 종 동정에서 진균류 분류에 주로 이용되는 유전자는 rRNA 유전자이다 (Harmesen *et al.*, 1995). rRNA 유전자 염기서열을 기반으로 지구상 생물체를 크게 박테리아, 원핵생물, 진핵생물로 구분할 수 있다. 또한 rRNA 유전자는 모든 세포 내에 존재하며 생명체의 단백질을 생성하는 필수적인 구성 요소로 진화상 다른 조상을 가지는 생명체들 간의 차이점을 설명할 수 있다. rRNA 유전자의 LSU(Large Subunit)와 SSU(Small Subunit) 부분은 서열의 변화가 작아 보통 속(Genus) 이상의 분류체계를 결정하기 위한 목적으로 많이 이용되며 ITS(Internal Transcribed Spacer)는 서열의 변화가 많이 일어나 종(Species) 수준의 동정을 하기 위해 많이 사용된다. rRNA 유전자 이외에도 진균류 동정 및 분류에 있어 이용되는 유전자로는 *cox1*, *elongation factor1- α* (*efl- α*), RNA polymerase II subunit(*rpb1*, *rpb2*), *tefl*, *atp6*, β -tub 등이 있다(Hibbett *et al.*, 2007).

하지만 현재까지 밝혀진 유전자 마커로는 진균류의 정확한 분류에 한계를 가진다. rRNA 유전자는 진균류의 경우 분류할 수 있는 단계의 범위가

그룹군마다 다른 경우가 많아 *Penicillium*에서는 같은 속(Genera), 다른 종이라도 ITS와 LSU 부위가 100% 동일해서 형태학적 방법을 병행해야 한다. 또한 *c.acremonium* 과 같이 동일한 속(Genera)에 속한 종이라도 5-10% 까지 유전자의 염기서열이 달라지기도 한다(Jim *et al.*, 2006). 그리고 진균류에서 60% 이상을 차지하고 있는 자낭균류 중 몇 속(Genera)에서 ITS 서열 부위의 다양성이 나타나지 않아 종의 명확한 분류를 할 수 없다. 뿐만 아니라 ITS 부위 길이의 차이와 시작 서열 부위의 다양성으로 인해 ITS Primer에 따른 오류가 생겨 Sequencing 시 부정확한 서열 정보를 얻을 수 있다. Cox1 유전자는 미토콘드리아에 존재하는 유전자 서열로 인트론 부위의 이동이 잦아 정확한 분류를 하기 어렵고 서열상 보존된 부위가 부족하여 PCR에 필요한 Universal Primer 개발이 어렵다는 단점을 가진다(Seifert *et al.*, 2009). 그리고 Efl- α 유전자는 세부 종 분류에 적용하는데 있어 서열 차이가 부족하여 체계적인 분류를 할 수 없다(Andrew *et al.*, 1999). β -Tub 유전자는 유전체 안에서 Multiple Copy Gene으로 존재하기 때문에 서열이 다른 부위로 인식 될 수 있다(Landvik *et al.*, 2001). Rpb1, Rpb2 유전자는 ITS 부위보다 높은 다양성을 가져 Species 분류에 이용되지만 *Cortinarius* 속(Genus)에서는 낮은 변별력을 가져 분류가 어렵다(Frøslev *et al.*, 2005). 이러한 단일 유전자 마커의 단점을 최소화 하고자 여러 개의 유전자를 이용하는 Multi-gene Phylogeny를 시도하기도 한다. 균류학자 사이 강한 네트워크 기반을 형성하고 있는 Assembling the Fungal Tree of Life(AFTOL) Project에서는 6개의 유전자(LSU, SSU, rpb1, rpb2, tef1, mtSSU) 마커의 조합을 이용하여 195 taxa(Order)를 대상으로 계통수를 분류하였다. 이를 바탕으로 *Ascomycota*, *Basidiomycota*, *Microsporidia*, *Chytridiomycota*, *Blastocladiomycota*, *Neocallimastigomycota*, *Glomeromycota*로 진균류를 7개의 문(Phylum)으로 나누었으며 현재 균류의 분류체계로 받아들여지고 있다(Hibbett *et al.*, 2007).

이와 같이 형태학적 분류의 한계로 유전자 서열을 이용한 진균류의 분류의 중요성이 커지고 있다. 하지만 종의 다양성이 매우 높은 진균류와 같은 경우 단일 유전자 마커를 이용하여 종, 속, 과, 목, 강, 문, 계 의 분류체계에 맞추어 체계적으로 분류하는 데에는 아직 한계를 가지며 이러한 단일 유전자 마커의 한계를 극복하기 위해서는 **Genome** 서열 분석을 바탕으로 새로운 유전자 마커의 발견 및 여러 유전자 마커를 조합하여 분류체계의 기반을 마련해야 할 것이다.

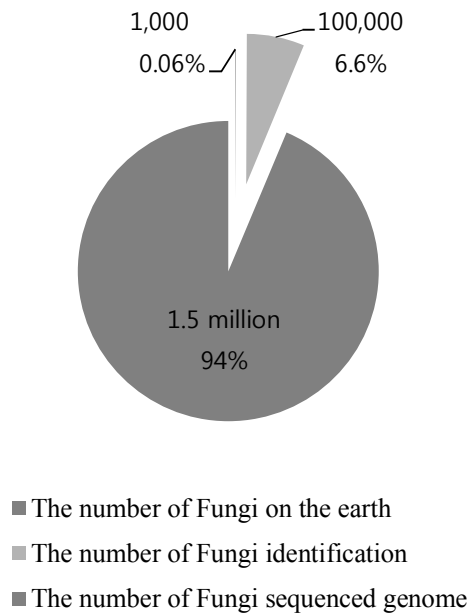


Figure 1.1 The total number of fungi present in the world. It has been predicted that the number is about 1.5 million, among them only about 6% of them have been known (total of 100,000 species).

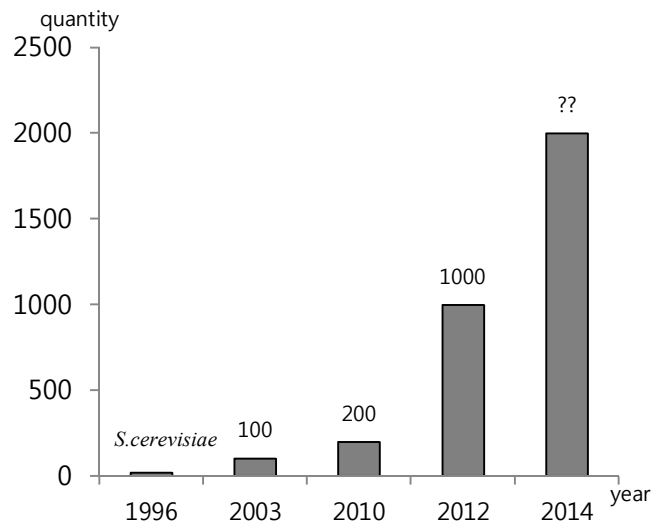


Figure 1.2 The total number of fungi genome sequence completed. Since the genome sequence of *S.cerevisiae* was revealed for the first time in 1996, the number has been increasing rapidly.

1.2 진균류 유전정보 관련 데이터베이스

진균류의 수 많은 유전자 서열 정보와 지놈 데이터의 축적으로 인해 데이터를 효율적으로 저장 및 이용할 수 있는 데이터베이스가 필요하다. NCBI의 GenBank는 전 세계에서 산출되는 모든 생물체의 유전자 서열 데이터가 저장되어 있는 최대의 생물정보 데이터베이스이다. GenBank에는 진균류 약 10만 종의 유전자 서열이 저장되어 있으며 80% 이상이 rRNA 유전자 서열이다(Hibbett *et al.*, 2007). 다양한 진균 종이 존재하는 만큼 Phylum(문)에 따라 독자적인 데이터베이스와 진균류의 전체 지놈 서열을 저장해 놓은 데이터베이스가 국내외 구축되어 있다. <Table 1.1>에 Fungi와 관련하여 구축된 국내외 데이터베이스를 정리하였다. 진핵생물 중 최초로 Genome서열이 밝혀진 *S.cerevisiae*의 Genome 정보와 생물학적 통합 정보를 제공하는 SGD(*Saccharomyces* Genome Database)는 유전자의 서열과 이들이 갖는 기능을 알 수 있다(Cherry *et al.*, 1998). 1998년에 구축된 이 데이터베이스는 현재까지도 많은 균류학자들에게 유용한 정보를 제공해 주며 유전자 서열의 상동성 검색 및 유전자 온톨로지를 이용하여 어노테이션이 가능하다(Dwight *et al.*, 2002). FGSC (Fungal Genetic Stock Center)에서 제공하는 데이터베이스는 *Neurospora*, *Aspergillus*, *Sordaria* 등 특정 진균류의 지놈 정보를 제공하며 GenBank와 연동하여 진균류 동정에 이용할 수 있다. 2009년까지 7만5천 개의 Strain의 유전자 데이터가 저장되어 있고 분산된 모든 사상균류 데이터를 포함하고 있다(Kevin *et al.*, 2009).

진균류의 생활사는 환경에 따른 영향을 많이 받기 때문에 지역에 따라 분포하는 종의 차이를 보인다. 이 때문에 국가 또는 대륙 별로 얻어진 진균류를 대상으로 데이터베이스를 구축하기도 한다. Pacific Northwest Fungi Database는 태평양 연안 북서부의 진균류를 수집하여 서열 정보를 얻고 분류체계를 정립하였다. 또한 진균류가 감염될 수 있는 Host의 정보와 서

식지의 특성을 함께 제공한다(Frank *et al.*, 2007). 오스트리아에 서식하는 진균류의 정보를 모아 놓은 Austrian Fungi Database는 7천여 곳 이상의 지역에서 얻은 31만개의 데이터를 저장하고 있다. 유럽에서 구축한 Q-bank는 진균류 DNA 서열을 바탕으로 다양한 바이오인포매틱스 분석을 수행할 수 있으며 종 동정 및 분류가 가능한 바코드 서열 정보와 함께 생태학적 특징을 함께 알려주어 형태학적 특징과 유전자 서열 분석이 가능하다(Arzanlou *et al.*, 2012). UNITE 데이터베이스는 진균류 중 외균근(Ectomycorrhizal)의 rRNA 에 속한 ITS 부위의 서열을 통합하였다. ITS 부위는 진균류의 유용한 바코드 마커로 이용되고 있으며 현재까지 273 Genera(속) 1,514 종의 ITS 서열 정보를 가지고 있다. 또한 INSD(International Nucleotide Sequence Databases)에서부터 얻은 서열과 함께 독자적으로 구축한 ITS 서열 정보를 통합하여 쿼리 서열을 입력하면 상동성 검색을 통하여 가장 유전자 서열이 높은 종의 정보 및 Phylogenetic Tree 결과를 보여주어 ITS 부위를 이용한 진균류의 종 동정 및 분류가 쉽게 이루어 질 수 있다(Koljal *et al.*, 2004).

이 외에도 진균류의 EST 데이터 정보를 기반으로 한 FUNNYBASE는 진균류에서 공통적으로 가지고 있는 효소와 관련된 EST 서열로 가계도를 작성할 수 있다(Sylvain *et al.*, 2008). 진균증을 일으키는 곰팡이를 중심으로 구축한 데이터베이스도 많이 존재하는데 칸디다증을 일으키는 칸디다의 유전정보를 가지고 있는 Candida DB는 지놈 서열을 바탕으로 하여 유전자의 기능 및 유전자 산물인 효소, 분비 단백질 등에 대한 정보를 알 수 있다(d'Enfert *et al.*, 2005).

국내에 구축되어 있는 진균류 관련 데이터베이스 CFGP(Comparative Fungal Genome Platform)는 진균류 서열을 기반으로 독립적인 데이터베이스 저장 공간을 마련하고 바이오인포매틱스 기술을 이용하여 연구할 수 있는 다양한 tool을 웹 상으로 제공하고 있다(Choi *et al.*, 2010). 이와 같은 진균

류의 유전서열 데이터를 이용한 데이터베이스 구축을 위한 노력은 국외를 비롯해 국내에서도 활발히 이루어지고 있다. 하지만 진균류의 분류 및 종 동정을 위한 유전자 마커 서열을 통합한 데이터베이스는 부재한 실정이다. 특히 진균에 의한 감염사례가 증가하고 새로이 밝혀진 진균들의 유전서열 정보의 축적에 따라 많은 종을 분류할 수 있고 새로운 진균을 빠르게 동정할 수 있는 유전자 서열 기반의 데이터베이스 구축은 매우 중요하다.

Table 1.1 Database for fungal research. There are varieties of databases abroad based on diverse fungi and mycoses.

Database	Contents	Web Database Address
SGD	Genome of <i>Saccharomyces</i>	http://www.yeastgenome.org/
Candida DB	Candida Genome, genetic information	http://genolist.pasteur.fr/CandidaDB/
FGSC	Fungal Genetics Stock Center	http://www.fgsc.net
FUNYBASE	Fungi EST Data & Complete Genome	http://genome.jouy.inra.fr/funybase
US National Fungus Collections	Fungi Specimens, Nomenclature, Index of Fungi	http://nt.ars-grin.gov/fungalatabases/
Pacific Northwest Fungi Database	Pacific Northwest fungus perspective and Taxonomic for fungi or host	http://pnwfungi.wsu.edu/programs/aboutDatabase.asp
e-Fungi	Fungal data resource for comparative analysis of fungal	http:// www.e-fungi.org.uk .
Candida Genome Database	Candida albicans genomic sequence data, Protein information	http://www.candidagenome.org/
DFCI Fungi Gene Indices	Analysis of Fungi EST and Gene Sequence	http://compbio.dfci.harvard.edu/index.html
FSD	Integrated platform for annotation of fungal secretomes	http://fsd.snu.ac.kr/

1.3 진균에 의한 감염성 질병

진균류 중에는 버섯, 누룩 등의 식이재료와 페니실린 등 의약품의 원료로 인간의 생활에 도움을 주기도 하지만 병원성을 일으키는 곰팡이균은 사람이나 동물에 기생하며 질병을 유발하고 식물에게 감염되어 식물의 생장 및 번식을 막는 등 많은 피해를 준다. 진균류가 일으키는 질병은 인간의 건강에 약하게는 알레르기성 질환, 피부염, 기관지 염에서 심하게는 곰팡이성 폐렴을 유발하여 사망에까지 이르게 한다. 지난 10년간 곰팡이를 포함한 미생물 감염 사례가 병원 내 면역 결함 환자에게서 더욱 심각히 보고 되었는데(Justana *et al.*, 2009) 이러한 감염성 질병은 인구집단의 크기가 증가 함에 따라 위험에 노출되는 집단도 증가하여 발생빈도가 급격히 증가되고 있음을 뜻한다(Richardson *et al.*, 2005)

모든 진균들 중 약 150개의 진균들이 인간에게 병원균으로 작용하게 되는데 곰팡이성 감염으로 보고된 종은 주로 *Candida*, *Cryptococcus*, *Aspergillus*이다(Chandler F, 1996). 하지만 이 외에도 사람에게 병을 유발하는 진균류는 매우 다양하고 환경에 따라 다르게 나타나며 빠른 시간 내에 다른 곳으로 전파할 수 있는 능력을 가지고 있기 때문에 간과할 수 없다. 이러한 진균 감염성 질병들은 발생 초기에는 진단이 어려울 뿐만 아니라 효과적으로 치료하는데 어려움이 있다(Cruz *et al.*, 2011). 왜냐하면 곰팡이는 생육할 수 있는 조건이 갖춰지면 단시간 내에 급속도로 증식하며 증식에 필요한 균사 또는 포자를 생성하여 공기 중에 병원균을 퍼지도록 하여 초기 진균을 잡아 내기 어렵기 때문이다. 또한 체내에 부작용을 일으키지 않는 효과적인 항진균제의 개발이 아직 미흡한 실정이다(Sharon *et al.*, 2010). 곰팡이 감염균은 매년 새로운 종이 보고되고 있으며 이로 인한 감염사례 또한 증가하고 있어 새로운 종의 동정과 효과적인 항진균 치료제의 개발이 시급하다. 또한 의료기술의 발달로 인한 장기이식과 항암치료제 그리

고 고령화의 시대로 접어들어 따라 많은 환자들이 면역력 저하로 이차 병원성 곰팡이 균 감염에 대비가 필요하다. 장기 이식을 한 환자나 항암치료를 받은 환자에게서 진균 감염의 빈도는 바이러스 감염의 빈도보다 적으나 치사율은 매우 높다. 이를 위한 병원균의 특성과 새로운 종의 발견 시 종의 동정을 위해 진균류의 분류과정은 필수적이다. 보건학적으로 곰팡이 질환에 대한 연구와 분류 기술의 연구는 인간이 더욱 건강한 삶을 살도록 하기 위한 필수적인 연구분야이다. 진균증(Mycoses)은 인간을 포함한 동물에게 진균으로 인해 나타나는 질병을 통칭하는 용어이다. 진균증은 숙주에게 침입하는 과정과 병독성에 따라 나뉘게 된다. 피부나 모발, 손톱, 발톱을 통한 감염은 표재성 진균증(Superficial Mycoses) 이라고 하며 피부 아래의 피하조직에 침입을 하는 피하 진균증(Subcutaneous Mycoses), 내부장기를 통한 감염인 전신성 진균증(Systemic Mycoses), 마지막으로 면역시스템에 문제가 있는 사람에게 일어나는 기회성 진균증 (Opportunistic Mycoses)이 있다(Cruz *et al.*, 2011).

주로 피부로 침투하는 표재성 진균증(Superficial Mycoses)은 동물에게 감염 될 수 있으며 사람에게까지 전달될 수 있다. 표재성 진균증은 세계 전체의 인구에서 20-25%의 높은 전파력을 보인다(Blanka *et al.*, 2008). 또한 지속적인 습도와 따뜻한 온도가 유지되는 여름과 열대 지방에 사는 사람에게서 자주 관찰 된다(Maksymiuk *et al.*, 1984). 이는 주로 피부사상균(Dermatophytes), 효모(Yeast) 등에 의해 일어나며 백선, 전풍, 피부 칸디다증과 같은 질병을 야기 시키며 피부의 각질층, 체모 및 손톱과 같은 케라틴에 기생하고 번식하면서 피부병을 일으킨다(Elewski *et al.*, 1989). 잘 알려진 질병으로 무좀, 아토피 피부염 등이 있다. 진균증 피부사상균의 종류로는 *Trichophyton*, *Microsporum*, *Epidermophyton* 등이 있다(Maksymiuk *et al.*, 1984). 피부사상균은 각질을 분해할 수 있는 능력을 가지고 있어 다른 세균이나 진균보다 각질을 선호하여 사람에게 상처 난 피부를 통해 감염을

일으키게 된다(Heiner, 2003). 백선의 원인균은 *T.rubrum*이 현재까지 가장 많이 동정된 균으로 감염되면 가려움증, 버짐 등을 일으킨다. 백선 중 가장 감염률이 높은 족부백선은 성인에게서 주로 나타나며 두부백선(*Tinea capitis*)는 아이들에게 많이 발생한다. 이를 일으키는 주요 종으로는 *M.canis*가 있으며 *T.tonsurans*로 인한 감염 또한 최근 꾸준히 증가하고 있다(Ameen, 2010).

면역저하 환자에서의 진균 기회감염은 주로 장기 이식을 받은 환자, 항암치료를 받은 환자에게 면역기능을 억제하는 치료요법을 사용하여 항체의 생성과 면역반응이 약화되었을 때 일어나는데 진균이 감염되었을 때, 정상인에 비해 감염이 되기 쉽고 반응속도 및 전이가 빨라져 악영향을 미치게 된다. 피부 칸디다증은 피부를 통한 *Candida*의 감염으로 발생하는 증상으로 통풍이 잘 되지 않고 습한 피부 부위에서 체액이 스며나오며 인접 부위에 작은 농포가 동반된다(Lachke *et al.*, 2003). 이러한 표재성 진균증의 진단은 주로 외부에 감염 증상이 나타나야 가능하며 또한 두부백선의 경우 일반적인 의학진단 방법은 감염된 환자들에 따라 다른 부정확한 측정을 보였다. 최근에는 분자적 진단방법이 시행되어 유전자를 기반으로 한 방법 등이 도입 되어 진단시간을 단축시키고 정확한 진단을 할 수 있게 되었다.

피하진균증 (Subcutaneous Mycoses)은 피부 바로 아래 조직인 피하지방, 근육으로 병원균이 침투하여 사람에게 감염되는 질병이며 사마귀 모양의 부스럼이 생기는 것이 특징이다. 이는 주로 오염된 식물의 가시, 유리조각 등을 통해 감염되며 숙주세포의 환경에 적응하여 만성적이고 감염 진행속도가 느리다(Telsuya *et al.*, 2003). 진균증의 종류로는 *Sporotrichosis*가 가장 흔하며, 최근에는 *Dematiaceous*에 의한 감염증인 색소분아진균증 (*chromoblastomycosis*), 흑색진균증(*phaeohyphomycosis*) 및 진균종(*eumycotic mycetoma*)이 증가하는 추세이다. 그 외 피부 효모균증

(*cutaneous cryptococcosis*), 피부 아스페르길루스증(*cutaneous aspergillosis*), 피부 모균증(*cutaneous mucormycosis*) 등이 보고되었다(Flavio *et al.*, 2003). 스포로트리쿰증(*sporotrichosis*)은 이형의 균류로 대부분의 국가에서 보고되었다. 이 균류는 따뜻하고 습한 지역의 죽은 식물에서 기생하는 기생식물에서 주로 발생하며 성별, 나이, 인종에 관계없이 대부분의 사람에게 감염시킬 수 있다(Kauffman *et al.*, 2006). 이 진균의 초기 발견은 1998년 Schenck에 의해 보고되었으며 현재에는 *Sporothrix schenckii*로 명명되고 있다(Schenck *et al.*, 1998). 감염경로는 주로 식물의 가시, 풀잎, 물이끼 등에 접촉에 의해 인체로 침입 할 수 있다. 이로 인해 감염계층은 주로 도시보다는 외곽 지역에서 식물과 많이 접하게 되는 계층에서 많이 발생된다. 색소분자진균증(*chromoblastomycosis*)은 darkly pigmented mold의 원인균이라고 알려져 있다(Flavio *et al.*, 2003). 감염초기에는 홍색의 발진으로 시작하여 점차 융기성 사마귀양 결절로 진행되는 경우가 많고 주로 30세에서 50세 사이의 남성들에게 발생한다(Ismael, 1989).

진균 감염 시 전신으로 퍼져 중증 질환인 전신성 진균증(Systemic Mycoses)은 호흡기를 통해 폐로 전달되거나 정맥혈을 따라 내부 다른 장기에 감염된다. 주로 초기에 폐렴을 일으키며 전신으로 퍼져 나가게 된다. 이러한 감염은 주로 흙에 사는 곰팡이에 의해 야기되며 공기 중에 날아다니는 포자를 호흡하여 감염이 시작된다. 이러한 전신성 진균증은 감염 시 빠른 속도로 혈액을 통하여 다른 장기에까지 감염시키기 때문에 치료가 쉽지 않다(Blanka *et al.*, 2008). 전신성 진균감염을 유발하는 진균종으로는 *Coccidioides immitis*, *Histoplasma capsulatum*, *Blastomyces dermatitidis*, *Paracoccidioides brasiliensis* 등이 있다(Fraser *et al.*, 1979).

기회감염에 의해 발생하는 기회진균증(Opportunistic Mycoses)의 경우는 장기이식 후 면역 억제제를 투여 받는 환자, 후천성 면역결핍증후군 환자(AIDS), 항균제, 스테로이드를 투여 받는 환자에서 주로 나타나며 심하

면 사망에까지 이르는 심각한 결과를 초래하게 된다. *Aspergillosis*는 대다수의 건강한 사람에게 노출 되었을 때에는 감염을 일으키지 않지만 면역 기능에 문제가 있는 사람에게서 질병을 일으켜 폐와 같은 호흡기 계통에 일차 감염을 일으키는데 안구 및 각막염, 피부, 중추신경계 간, 신장, 심장, 뇌, 간 등 신체 내부 기관에까지도 감염이 일어나게 된다. *Mucormycosis*은 털곰팡이증 이라고 하며 보통 당뇨병의 합병증으로 일어나게 된다. 감염 경로는 주로 상부기도나 허파에서 시작되는 경우가 많고 심한 경우 출혈 폐렴, 혈전부위 경색을 일으키게 된다. 효모(Yeast)에 의한 기회감염의 경우 혈액성 칸디다증이 있다. 이는 항암치료로 인해 면역력에 약해져 있는 암환자들에게 주로 나타난다(Anaissie, 1992). 진균증은 이처럼 사람에게 약하게는 가려움증, 홍반 등 에서부터 심하게는 전신으로 퍼져 사망에 이르게 한다. 이에 따라 정확한 진균증 진단 및 빠른 치료기법이 개발되어야 할 것이다. <Table 1.2>에서는 진균증에 따른 관련 진균 종 명과 진균증 증상들을 정리하였다.

많은 인구가 현재에도 진균증에 고통 받고 있으며 이에 따른 치료법 또한 많이 생겨나게 되었다. 하지만 진균증 감염의 초기 진단이 어렵고 외재적 특징이 나타나기 시작하면 이미 진균증은 신체의 전과경로를 따라 많은 곳으로 전이되거나 더욱 심각해져 완벽한 치료를 하는데 시간이 오래 걸리고 치료 효과가 다르게 나타날 경우 심하게는 급성으로 사망하게 된다. 특히 자가치유 능력이 약해져 있고 신체 대사 흐름이 저하된 고령의 환자인 경우에는 한번 진균에 감염이 되면 사망에 이르기 쉽다. 이러한 문제점을 해결하기 위해서는 보건학적으로 인구 집단에 대한 접근이 필요며 진균증은 공기 중이나 접촉을 통하여 많은 사람에게 전파 될 수 있어 초기 예방이 더욱 중요하다. 더욱이 장기입원 환자의 경우 오랜 치료로 인해 면역력이 약해져 있기 때문에 진균에 의한 2차 감염 예방을 병원 측에서는 더욱 신경을 써야 한다. 표재성 진균증의 경우 피부에 발진

이 생기는 등 생명에는 치명적이지는 않지만 세균 감염증 보다 흔하지 않아 과거 항생제에 비해 항진균제의 개발이 활발하지 못했다. 하지만 면역억제제의 사용으로 인해 표재성 진균증의 감염이 과거에 비해 증가함에 따라 강력한 항진균제의 개발이 필요하다. 또한 이러한 감염성 곰팡이의 경우 세대를 거듭함에 따라 진화적으로 유전적 전이 및 서열의 변화를 일으켜 감염력 또한 다르게 나타나게 된다. 새로운 진균류의 발견 시 감염 및 전파여부를 알기 위해서는 유전정보에 기반한 독성 연구가 필요하다. 또한 숙주와의 유전자 상호작용은 숙주를 사멸하게 하는 결과를 초래하게 되므로 진균류의 유전자 서열에 기반한 데이터베이스 마련 및 분석이 요구된다.

Table 1.2 Symptom of mycoses. Classification of mycoses with virulence and symptom.

Mycoses	Fungi	Symptom
Superficial Mycoses	<i>Trichophyton</i> , <i>Microsporum</i> <i>Epidermophyton</i> ,	athlete's foot, ringworm, jock itch tinea, keratinized tissue tinea corporis, tinea cruris, tinea pedis athlete's foot, ringworm, jock itch
Subcutaneous Mycoses	<i>Sporotrichosis</i> <i>Chromoblastomycosis</i> <i>Cutaneous aspergillosis</i> <i>phaeohyphomycosis</i> .	nodule, ulcer, infect bone, lung, brain small red papule, lesion papule, erythema, ulcer, nodule dark bloom, cyst, trauma
Systemic Mycoses	<i>Blastomycosis</i> <i>Cryptococcosis</i> <i>Zygomycosis</i> <i>Mucormycosis</i>	diarrhea, bloating, discomfort, flatulence, constipation, bronchopulmonary disorders, asthma, breathing difficulties, fatigue, allergies, weight loss, fever, chills, malaise, depression
Opportunistic Mycoses	<i>Aspergillosis</i> <i>Mucormycosis</i> <i>Candidiasis</i> <i>Trichosporonosis</i>	cough, fever, jaundice, breathing difficulties thrombosis, loss of blood supply, headache itching, burning, irritation fever, high mortality

1.4 진균의 계통수

계통수(Phylogenetic Tree)는 생물 군의 표현형이나 유전적 특징을 바탕으로 공통조상으로부터 진화되어 파생되어온 분류 군 사이의 유연 관계를 보여줄 수 있는 나뭇가지 모양의 그림이다. 계통수를 작성하는 이유는 한번에 분류 군 사이의 상호 연관 성 및 공통 조상에 대한 정보를 얻기 쉽기 때문이다(Fitz *et al.*, 1999). 뿐만 아니라 종 분화 시점 및 진화 과정을 설명할 수 있으며 이미 기능이 알려진 유전자, 단백질 서열과의 상동성 비교를 통해 새로 발견한 생물 종 및 유전자의 기능을 예측할 수 있다. 현재 지구상에 존재하고 있는 생물 종의 계통분류 체계는 크게 세균, 고세균, 진핵생물 이렇게 3개로 나뉘어 진다. 균계는 원생생물, 식물, 동물이 있는 진핵생물에 포함되어 있다(Lorenzo *et al.*, 2004). 균계가 진핵생물로서 가지는 특징은 핵이 핵막으로 싸여 있으며 세포질에 미토콘드리아와 리보솜이 존재하고 있다. 또한 유전자에 인트론, 엑손 부위가 포함된 DNA 서열을 가지고 있다. 진균류는 동물과 식물 각각의 특징을 가지고 있어 따로 균계로 분류되어 나오게 되었는데 동물의 특징과 유사한 것은 엽록체가 없어 광합성을 하지 않아 스스로 양분을 만들지 못해 숙주에 기생하여 생활하는 것이다. 식물로서 가지는 특징은 식물세포의 세포벽처럼 진균류도 chitin으로 구성된 세포벽으로 둘러 싸여 있으며 유성생식, 무성생식을 하며 포자를 만들어 번식을 하는 것이다. 진균류가 가지는 이러한 형태학적, 생활사적 특징들에 기반한 계통분류는 rRNA 유전자 서열을 기반으로 했을 때도 유사한 분류체계 결과를 보인다(Baldauf *et al.*, 1993).

과거 진균류의 계통수는 생활 주기를 비롯하여 배우자, 배우자낭, 포자 등의 구조적, 세포학적 특징에 따라 병꼴균문, 자낭균문, 접합균문, 담자균문, 불완전균문 이렇게 5개의 문으로 나뉘어 졌다(Leedale *et al.*, 1974). 하지만 현재에는 유전학, 분자생물학의 발달로 인해 진균류의 유전체 서

열이 알려지면서 유전자 서열을 바탕으로 분류체계가 이루어 지고 있다. 유전자 서열을 기반으로 한 계통수 작성은 종을 대표할 수 있는 유전자 마커가 필요하다. 유전자 마커로 이용되기 위한 조건으로는 연구하고자 하는 마커 유전자가 모든 종에 존재해야 하며 중간 진화과정을 거치는 동안 돌연변이로 인한 유전자 서열의 결손, 치환, 삽입 등이 일어나야 한다. 또한 각 분류학적 위치에 따라서 서열 변이의 차이를 보여야 하는데 종간의 계통수를 구분하기 위해서는 서열의 변이가 비교적 많아야 한다 (Bazin *et al.*, 2012). 현재까지 진균류의 계통수를 작성하기 위해 다양한 유전자 마커가 이용되었으며 2007년 진균류의 정확한 계통수를 작성하기 위해 AFTOL(Assembling the Fungal Tree of Life) 프로젝트가 진행 되었다. 진균류의 6개의 유전자 서열을 바탕으로 Higher-level Phylogenetic Classification을 수행 한 분류에서 최종적으로 7개의 문(Phylum)으로 분류 하였다. 여기에서 분류된 7개의 문(Phylum)은 자낭균문(*Ascomycota*), 담자균문(*Basidiomycota*), 호상균문(*Chytridiomycota*), 취균문(*Glomeromycota*), 미포자충(*Microsporidia*), 네오칼리마스트릭스균문(*Neocallimastigomycota*), 블라스토클리디아균문(*Blastocladiomycota*) 이다(Hibbett *et al.*, 2007). <Figure 1.3>은 AFTOL 에서 유전자 서열에 기반하여 7개의 Phylum에 따른 진균류 분류 체계를 나타낸 것이다.

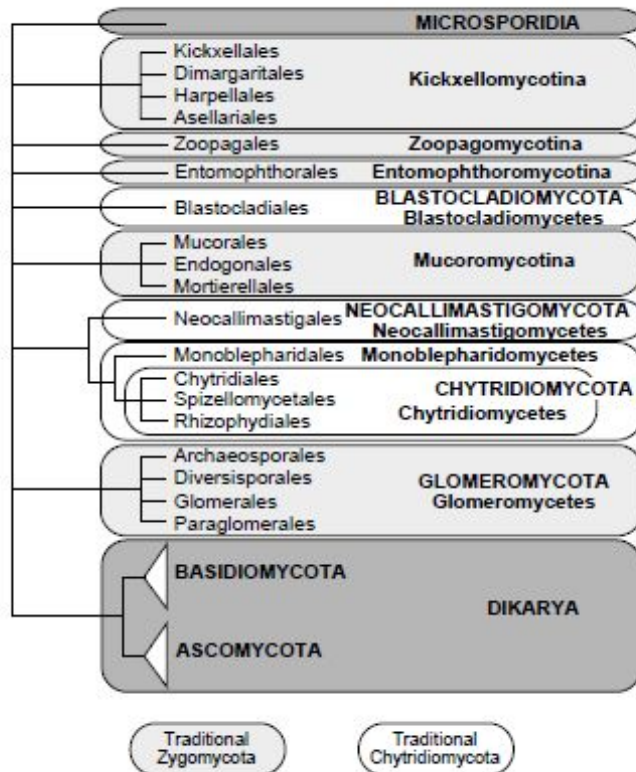


Figure 1.3 Seven phylums based on AFTOL project. In AFTOL, the fungus is divided into seven phylums based on genetic markers. (Hibbett *et al.*, 2007)

현재까지 종 동정이 완료된 진균류는 자낭균류가 약 6만4천 종으로 가장 많이 존재하며 그 다음으로 담자균류로 약 3만종 정도가 있다. 미포자충의 경우 과거에는 원생동물로 분류 되었는데 유전자 서열을 바탕으로 계통수를 작성하면 접합균류와 진화적으로 유사한 결과를 나타낸다 (Tanabe *et al.*, 2002). 이렇게 형태학적 분류체계는 유전학적 분류체계로 인해 차츰 변해가고 있으며 이에 따라 유전자 서열을 기반으로 한 계통수는 더욱 중요하다. 계통수는 공통조상에서부터 나와 Clade를 형성하게 되는데 정확한 계통수는 종의 분류체계에 따라 적합한 Clade를 형성해야 한다. 진균류는 린네의 분류체계를 적용하여 종, 속, 과, 목, 강, 문, 계 체계로 분류된다. 각 문(Phylum, 門)의 명칭이 -mycota 로 끝나며 문 이하의 분류군으로 아문(SubPhylum, 亞門)에는 -mycotina, 강(Class. 綱)에는 -mycetes, 아강(SubClass. 亞綱)에는 -mycetidea, 목(Order, 目)에는 -ales, 과(Family, 科)에는 -acea 라는 어미가 붙으며 마지막 속명과 종명은 이명법을 도입하여 속명 뒤에 종명이 나오게 된다(Deacon, 2006). <Figure 1.4>는 최초로 지놈 서열이 밝혀진 효모에 속하는 *S.cerevisiae* 에 대한 분류학적 위치를 나타내었다. 이처럼 각 진균종은 형태학적 특징과 유전자 서열을 기반으로 분류체계가 정립되어 가고 있으며 각각의 분류체계 단계별로 공통적인 특징을 갖고 있다. 진균류의 분류체계 정립은 새로운 진균 종의 등장 시 빠르게 분류가 가능하며 질병을 일으키는 진균의 특성을 파악하는데 있어 기초가 될 것이다.

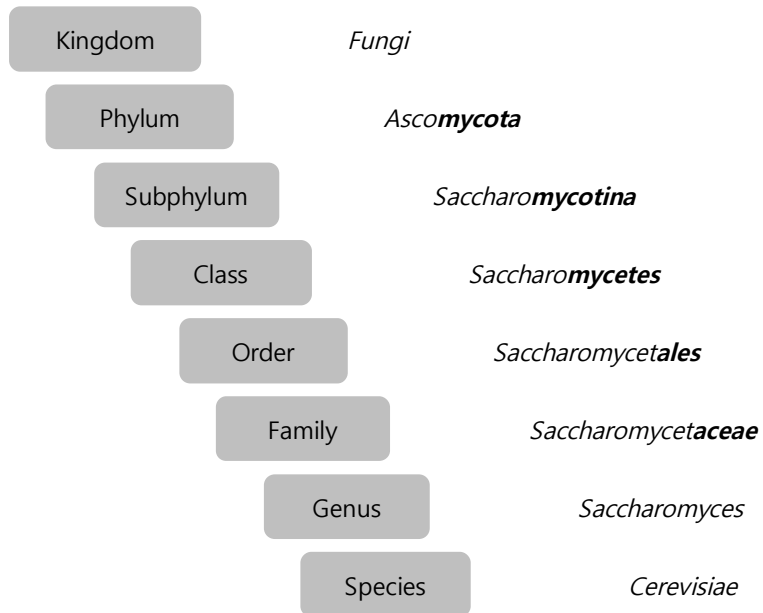


Figure 1.4 Classification of *S.cerevisiae*. *S.cerevisiae* is classified in kingdom of fungi, phylum of Ascomycota, subphylum of saccharomycotina, classes of saccharomycetes, order of saccharomycetales, family of saccharomycetacea and genus of saccharomyces.

2. 연구의 필요성

진균류는 문명의 시작과 더불어 인간과 함께 공존하며 먹거리로의 이용뿐만 아니라 생태계의 분해자 또는 의약품의 원료로 이용되어 왔다. 반면 진균류는 다양한 생명체에 기생하여 생활하기 때문에 인간에게 감염되면 병을 유발하거나 심한 경우 죽음에 이르게 한다. 현재까지 150종 이상의 균류가 척추동물 및 인간에게 질병을 유발하는 것으로 알려져 있다. 최근 노령인구의 증가 및 장기 이식 후의 면역 억제제의 사용과 면역결핍 증으로 인한 기회감염이 증가하고 있다. 이러한 진균증은 장기간 치료를 필요로 하며 재감염 빈도가 높아 이를 막기 위한 지속적인 연구가 필요하다. 특히 전염성 무좀은 전체 인구의 15%정도가 감염되었으며 집단 생활을 통해 잘 전염이 된다. 하지만 이를 치료하기 위한 치료제들은 진균 감염을 완치하기 보다는 증상 완화를 위한 의약품이 주를 이룬다. 또한 생물 멸종을 일으키는 원인 중의 하나인 곰팡이 병원균이 전체의 70%를 차지한다는 점에서 진균에 대한 연구는 보건학적으로 매우 중요하다 (Matthew *et al.*, 2012).

전 세계에 예측되는 진균류의 숫자는 150만개이지만 현재까지 알려진 종은 7%에 불과한 약 10만개 정도이다. 이 중 전체 지놈 서열이 모두 밝혀진 진균은 200여종 밖에 되지 않는다. 최근 다량의 서열정보를 빠른 시간 내에 얻을 수 있는 NGS(Next Generation Sequencing)기법의 발달로 진균류의 유전 서열 정보가 급속도로 증가 될 것이며 앞으로 더욱 기하급수적으로 방대한 진균류의 유전학적 데이터가 쏟아져 나올 것이라 예상된다. 이에 따라 진균류의 계통 분류학적 체계의 정립은 필수적이며 이를 분류할 수 있는 유전자 마커의 발견 또한 절실히 요구 된다. 방대한 유전서열 데이터는 계속 축적될 것이며 이를 효율적으로 이용하는 것뿐 만 아니라 새로운 진균의 계통분류를 위해서는 지금까지 밝혀진 진균류의 유전 데이

터의 통합이 필요하다. 이미 많은 연구에서 진균류의 유전 서열 데이터를 이용하여 Phylogenetic Tree의 작성과 BLAST를 이용한 상동성 비교 등 다양한 연구 결과를 얻고 있다. 하지만 진균류의 진화과정의 복잡성과 종의 다양성으로 인해 유전 서열을 기반으로 한 분류체계에 한계가 있어 진균류 동정을 위한 신뢰성 있는 기준이 필요하다. 현재까지 알려진 진균류의 유전자 마커는 약 20여가지 이다(Nicolas *et al.*, 2009). 하지만 이러한 단일 유전자 마커는 균류의 계통분류에 한계가 있기 때문에 여러 유전자 조합을 이용한다. 가장 적합한 유전자 마커의 조합을 찾기 위해서는 각 종마다 유전자 마커의 서열 정보가 필요하며 이를 위해 원하는 유전자 서열을 간편하게 얻을 수 있는 데이터베이스 구축이 필요하다. 현재 많은 연구자들이 이용하고 있는 NCBI의 GenBank 데이터베이스는 종 이름 및 유전자명의 검색을 통하여 유전자 서열을 얻을 수 있다. 하지만 다양한 유전자 마커의 조합을 비교하기 위해서는 각 유전자의 정보를 포함하고 있는 종을 모르고 BLAST를 이용한 분석에서는 진균류 서열뿐만 아니라 모든 생물 종의 유전자 서열이 데이터베이스화 되어 있어 시간이 오래 걸리고 다른 종과의 서열 정보와 상동성 비교를 해야 하는 단점을 가진다. 각 국가의 진균류를 대상으로 한 유전자 서열 데이터의 통합과 분류체계의 마련은 균류에 의한 질병 예방 및 치료를 위한 연구의 바탕이 될 것이다. 현재 국내에는 전체 진균류의 유전서열 정보를 통합한 데이터베이스는 부재한 실정이다. 따라서 국내 진균 감염사례의 지속적인 증가와 진균증 치료를 위한 항진균제 및 치료기법의 연구를 위해서는 접근성이 용이한 국내 데이터베이스의 구축을 필요로 한다. 이를 바탕으로 다양한 진균류의 유전서열로 새로운 정보 창출이 이루어 질 것이며 산재된 데이터의 통합은 국내, 외 진균류의 분류체계 기반 마련에 기틀이 될 것이다.

3. 연구 목적

진균류의 유전 정보, 분류체계 정보 및 보건학적 데이터의 구축은 인간에게 감염성 질환을 일으키는 진균 관련 정보를 수집하여 보건 분야의 연구뿐만 아니라 진균의 계통분류 연구의 기반이 될 것이다. 이는 새로이 등장하는 감염성 진균의 빠른 종 동정과 진화적 위치를 확인하여 종에 따른 항 진균제 메커니즘을 빠르게 찾아 진균의 확산을 예방할 수 있을 것이다. 특히 향후 진균에 대한 진단 및 치료, 질병을 일으키는 다양한 진균의 분류, 지속적인 데이터의 보존 및 관리에 이용될 수 있을 것이다. 즉, 본 연구의 목적은 산재 되어 있는 전세계 진균류의 유전자 서열을 모두 모아 웹 상으로 데이터를 제공하고 서열의 상동성 검색을 통해 새로운 진균의 일부 서열로 계통 분류학적 위치를 빠르고 쉽게 알아낼 수 있도록 하고자 한다. 뿐만 아니라 보건학적 데이터인 진균증에 대한 정보를 포함하여 인간에게 해를 끼치는 진균류에 대한 연구의 바탕이 되고자 한다.

본 연구에서는 진균류에 유용하게 사용되는 유전자 마커 서열의 통합 및 진균류의 서열 데이터에 기반한 동정 및 종 분류를 위한 데이터베이스를 구축하였다. 이는 전산학적 기술을 이용하여 진균류의 유전자 서열을 핸들링하고 이에 기반한 데이터베이스를 웹 인터페이스를 통해 공개 함으로서 진균의 마커 유전자에 대한 정보 검색 및 서열 획득이 용이하도록 한 것이다. 또한 서열의 상동성 비교를 위한 바이오인포매틱스 툴을 웹 상으로 제공하여 유전자 서열 데이터를 이용하여 상동성 검색 및 종 동정이 가능하도록 하였으며 새로운 진균류의 동정 시 이용할 수 있는 분류기준을 마련하여 새로운 유전자 마커의 조합을 찾는데 유용한 데이터베이스를 구축하고자 하였다. 진균류의 유전자 서열은 INSD(International Nucleotide Sequence Databases)에 모두 저장되어 있으나 각 유전자에 따라 통합한 데이터베이스는 찾기 힘들다. 이에 본 연구에서는 유전자 별로 서

열이 밝혀진 종들의 분류체계를 함께 제공하여 진균류의 계통분류에 유용하게 이용될 수 있을 것이다. 특히 단일 유전자 마커의 한계로 여러 유전자를 조합한 **Multigene**을 이용한 계통분류의 필요성을 증가시키고 있다. **Multigene**을 이용하기 위해서는 조합하고자 하는 유전자들의 서열이 모두 데이터베이스화 되어야 한다. 하지만 아직 모든 종의 유전자 서열 정보가 밝혀지지 않은 시점에서 찾고자 하는 단일 유전자들의 종 정보를 통합하기에는 한계를 가진다. 이 때문에 본 연구에서는 저장되어 있는 유전자 서열 데이터 중 조합하고자 하는 유전자 서열의 종 정보를 제공해 준다. 또한 각각의 유전자 마커 서열의 통합은 새로운 진균 종에 대해 일부 유전자 서열 정보만을 가지고 기존의 데이터베이스 안에서 서열 상동성 비교를 통해 종의 분류체계 정보를 제공해 준다.

II. 연구방법

1. 데이터 수집 및 가공

본 연구에서 이용된 데이터는 GenBank(<ftp://ftp.ncbi.nih.gov/genbank/>)에서 수집 하였다. GenBank의 ftp 사이트에는 데이터로 구축되어 있는 모든 종의 서열이 저장되어 있는데 진균류는 gbpln 파일명으로 되어 있다. gbpln 파일명을 가진 데이터 안에는 진균류 서열뿐만 아니라 식물의 유전자 서열도 포함되어 있기 때문에 데이터 가공이 필요하다. 뿐만 아니라 데이터베이스에 필요한 정보만 추출하기 위해 JAVA 프로그래밍 언어를 이용하여 데이터 파싱 과정을 수행하였다.

gbpln_.seq 파일은 Fungi 뿐만 아니라 식물의 유전자 서열 정보를 포함 한 Genome 서열, Chromosome 서열, Contig 서열 등 많은 데이터들이 산재되어 있어 마커 유전자 정보만을 추출하기 위해 데이터 가공을 시행하였다. 데이터 가공은 프로그래밍 언어인 JAVA를 이용하였다. 또한 Fungi의 rRNA Gene, Calmodulin Gene, Elongation Factor1- α (EF1- α) Gene, RNA polymerase II subunit(RPB2) Gene, ATPase subunit 6 Gene, β -Tubulin Gene의 유전자 서열 중에서도 genome, chromosome, contig 서열을 제거하여 데이터를 재구성하였다. gbpln1.seq에서부터 gbpln59.seq의 flat file을 대상으로 데이터 추출 및 가공을 시행하였다. flat file에서 'ACCESSION' 행에서 서열의 고유 번호와 'SOURCE'에서 종의 이름, 'ORGANISM'에서 종이 속한 계통분류학적 정보, 'ORIGIN'에서 서열 정보만을 추출하였다. 이러한 데이터의 수집과정을 통해 BLAST에 기반이 되는 데이터베이스를 fasta 형식으로 재가공하기 위해 각 서열마다 '>' 표시를 시작으로 종 이름과 분류체계, Accession 넘버가 한 행 안에 모두 포함되게 한 후 아래 행에 유전

자 서열의 공백을 제거한 후 대문자로 바꾸어 70개씩 나열되도록 하였다. <Figure 2.1>과 <Figure 2.2>는 본 연구를 수행하기 위해 Genbank에서 저장한 데이터 파일과 이를 바탕으로 데이터 가공 후 형성된 데이터 형식을 나타내었다.

```

LOCUS      GU289393                      509 bp    DNA        linear    PLN 18-MAY-2010
DEFINITION Phomopsis sp. N 7.2 18S ribosomal RNA gene, partial sequence.
ACCESSION  GU289393
VERSION    GU289393.1  GI:296033468
KEYWORDS   .
SOURCE     Phomopsis sp. N 7.2
  ORGANISM Phomopsis sp. N 7.2
            Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina;
            Sordariomycetes; Sordariomycetidae; Diaporthales; Valsaceae;
            mitosporic Valsaceae; Phomopsis.
REFERENCE  1 (bases 1 to 509)
  AUTHORS  To,K.A., Nguyen,P.M. and Le,Q.H.
  TITLE    Isolation of laccase-producing Phomopsis sp. N 7.2
  JOURNAL  Unpublished
REFERENCE  2 (bases 1 to 509)
  AUTHORS  To,K.A., Nguyen,P.M. and Le,Q.H.
  TITLE    Direct Submission
  JOURNAL  Submitted (06-DEC-2009) Biochemistry & Molecular Biology, Hanoi
            University of Technology, Dai Co Viet, Hanoi, Hanoi 0000, Vietnam
FEATURES             Location/Qualifiers
     source            1..509
                       /organism="Phomopsis sp. N 7.2"
                       /mol_type="genomic DNA"
                       /strain="N 7.2"
                       /isolation_source="paper effluent sample"
                       /db_xref="taxon:707208"
                       /PCR_primers="fwd_seq: tcgtaggtgaacctgcgg, rev_seq:
                       tcctccgcttattgatatgc"
     rRNA              <1..>509
                       /product="18S ribosomal RNA"
ORIGIN
      1  tgctggaacg cgccctaggc gcacccagaa accctttgtg aacttatacc ttactgttgc
     61  ctcgggcgcat gctggccccc ttgggggtccc ctggagacag ggagcaggca cgccggcggc
    121  caagttaact cttgttttta cactgaaact ctgagaaaaa acacaaatga atcaaaactt
    181  tcaacaacgg atctcttggt tctggcatcg atgaagaacg cagcgaaatg cgataagtaa
    241  tgtgaattgc agaattcagt gaatcatcga atccttgaac gcacattgcg ccctctggta
    301  ttccggaggg catgcctgtt cgagcgtcat ttcaaccctc aagcattgct tgggtgtggg
    361  gcactgcctg taaaagggca ggccctgaaa tctagtggcg agctcgccag gaccccgagc
    421  gcagtagtta aaccctcgct ttggaaggcc ctggcgggtg cctgcccgtta aacccccaac
    481  ttttgaaaat ttgacctcgg atcaggtag
//

```

Figure 2.1 Genbank data before processing. It is a form of initial data obtained from the Genbank before carrying out this study. Gene sequences contain various kinds of information.

a)

```
>DQ218892 Geastrum sp. |Basidiomycota Agaricomycotina Agaricomycetes Phallomycetidae Geastrales
TTATTAGGAGATCTAAATTTAAGTTTAACTAATTTTCTTATATTCTATATTAGTATTATATTAGTGA
CAAGTATTCATCTATTAATTTAGAGGTCCTGACTTTATATCTAATAAATTCATAATAAATTAATTCCTTC
TTCTTGGAAATAGCTCTAGAGAGTTCTTACGCCAGTATAAATAGTATCGTGGAGAACAAATAGGTCCTA
AGAAATGAATATACTTACCTTTTATTTATTTCTTTTATATTAAATTCAAATTTAATAGGAA
ATACTCCATCTCTTTTACCATAACACTAGTATAATATTAGTATAGGATTTACTATTTTAT
AGGTGTGACAAATTTAGCATTATTTAAACACGGATTGCACCTTCTCTCTTTTTTATCCAGGGGGTACA
CCTTTAGGCTCTTGTCCACTGCTTTTAAAGAAAGTATAAGTTATTAGCTAGAGCTCTTTCTTTAG
GAGTTCGAATATTCGCCAATATGATAGCAGATCATACTTTATTAATAAATCTTGTCTACTTTCTATATCA
ATTATTACAGAGTGSAGTAAATGACTGCCATTTTACTCTAATACCTTTTCAATATTGTAGCTATTATT
GGTTTGAATAGCAGTTTCATTCAATCAA
>DQ218893 Myriostoma coliforme |Basidiomycota Agaricomycotina Agaricomycetes Phallomycetidae Geastrales
GCTCCTATATTAGGTTATTTAAATATAAGTTTAAACGAATTTTGCCCTATATACTATTATAGTTTATTT
TAGTAATAGCAATTCATATTTTATAGAGGTCCTCACCATTCTCTGAATCAATAGATCAACGTTTTCC
CTCTCTCATAGTAATTTAGTCTTCCACCTTGGAAATATAGCCTTAGAAAGTTCTTACGCAAGTATAAAT
AGTATCGTGGAGAACAAATAGGTATAAGAAATGAATTAATTTACCAATTTATTTATTCAATATCTTTT
TTATATCTGTGGCAAAATTAATAGGTAATCTCTTATCTTTTACCATACTACAGCATAGTTTATC
GGTAGGCTGAGTTTTTACCATAATTTATAGGTGTCACTTTAATTGCACTATTTAAACACGGACTCAATTT
TTCTCTTTTTTATTCAGGTGGACACCTTTAGSCCTTGGTTCCCTTGTAGTTTTAATAGAAATACAA
GTTATTGCTAGAGCTCTTTCTTTAGGGGTTGCAATTTTCTAATATGACAGCAGGTCTACTTTTAT
AAAAATTTTATCTACATCTTATATCAATTTATTTAGTAGTAGTATATTAATAGCTATCTTTACTTTAGT
CCATTGCAATCTTTTATGCAATTAATGGTTTGAATAGCAGTATCTATAATACAA
>DQ218894 Schenella simplex |Basidiomycota Agaricomycotina Agaricomycetes Phallomycetidae Geastrales
GCCCTCTCTTAGGTTATTTAAATATAAGTATAACAAATTTTCTACTATATCTATAATAGTGCTAATAT
TAGTGTGAATATCCACGTTTTATTTTCTCAAAATAGCGAGGAGAAAAGTAAGTTAACTCGCTCTCTTG
GAATATCGCCCTAGAAAGTTCTTATGCAAGCATAAATAGTATCGTGGGACAAATGGAATTAGGAAT
GAGATATATTTACCAATTTATTTATCTTTATTTTTTCAATTAATTTGCTAATTTAATAGGAATATAC
CCTATCTTTTACAATTACACTAGTATAATTTATCTGAGGTTAAGTTTACAATTTTATAGGTGT
TACTTTAATTGCAATTTTAAACATGGATTACACTTTTTTCTTTCTTTATACCAGGAGGTACACCTTTA
GCCTTAGTCCCTCTTTAGTTTTAATCGAAGTTACAGTTATCTAGCTAGAGCTTTATCTTTAGGAGTAC
GTTTATTGCTAACATGTGTGCAAGTCTACTCTCTTAAATATTATCAACATCTTATACCAATGTT
TAGTATTGGTTTATTAATGGCTGTATTACTTTAATTCCTTTTGAATATTGTGCACTATTGGGTTA
GAAATAGCCGTATCTATTATACAA
```

b)

EU339243	Antrodia carbonica	Basidiomycota Agaricomycotina Agaricomycetes	Polyporales GCACCTATTTTAGGATATTTAAATATAAGTT
EU339244	Auriporia aurea	Basidiomycota Agaricomycotina Agaricomycetes	Polyporales GTCCTATATTTGCGTTATATTAATATCAGTTTAAAC
EU339245	Bondarzewia montana	Basidiomycota Agaricomycotina Agaricomycetes	Polyporales AGTAATTTAGGTTTATCTACGATTTAGTTC
EU339246	Ceriporia purpurea	Basidiomycota Agaricomycotina Agaricomycetes	Polyporales GTCCTATTTTGGATACGTTCAATTTAAAT
EU339247	Cotylidia sp.	Basidiomycota Agaricomycotina Agaricomycetes	Polyporales GTCCTATTTTGGATATATAACATCACTCTCAAGTA
EU339248	Cryptoporus volvatus	Basidiomycota Agaricomycotina Agaricomycetes	Polyporales CTATATATGGATATTTAAATATTACATTA
EU339249	Dacrymyces chrysospermus	Basidiomycota Agaricomycotina Agaricomycetes	Polyporales ATACCATTATTAGGTTACATAAATCTT
EU339250	Dentocorticium sulphurellum	Basidiomycota Agaricomycotina Agaricomycetes	Polyporales GCACCTATTTTAGGTTATTTAAAA
EU339251	Echinodontium tinctorium	Basidiomycota Agaricomycotina Agaricomycetes	Polyporales GCACCTATTTTGGTTATTTAAAT
EU339252	Entoloma strictius	Basidiomycota Agaricomycotina Agaricomycetes	Agaricomycetidae Polyporales AAACCTAACCTTTAAG
EU339253	Ganoderma teugae	Basidiomycota Agaricomycotina Agaricomycetes	Polyporales ATATTAGGATATTTAAATATCTTTAACTAAT
EU339254	Gautieria othii	Basidiomycota Agaricomycotina Agaricomycetes	Phallomycetidae Polyporales GTCCTGTTTTTGGCTACT
EU339255	Heterobasidion annosum	Basidiomycota Agaricomycotina Agaricomycetes	Polyporales CCTATATTTGGTTATTTAAATTTAAGT
EU339256	Hyphodontia alutaria	Basidiomycota Agaricomycotina Agaricomycetes	Polyporales GTCCTATTTTGGTTATATCAACTTTAC
EU339257	Dendrocorticium roseocarpum	Basidiomycota Agaricomycotina Agaricomycetes	Polyporales AATATTTGGTTATTTTAAAT

Figure 2.2 Genbank data after processing using Java scripts. Genbank data are changed using java scripts a) fasta format for standalone BLAST b) sql format for table which has information of species classification.

2. 시스템 개발 환경 및 BLAST 서버 구축

진균류 동정을 위한 마커 유전자 기반 검색시스템 구축을 위해 HPC 클러스터 시스템을 기반으로 8C AMD Opteron-6128 2.0Ghz, 8Gb RAM, 500Gb SATA 7200rpm 3Gbps HDD 사양의 서버를 구축하였다. Operating System으로는 Linux v2.6.18을 사용하였으며 리눅스 서버 환경에서 데이터 저장을 위하여 MySQL을 이용하였다. 또한 데이터 파싱을 위해 JAVA 프로그래밍 언어와 웹과의 연동을 위한 프로그래밍 언어로는 JSP, HTML 그리고 자바 스크립트를 이용하였다. 웹 서버 프로그램으로는 Apache를 기반으로 하였으며 웹 컨테이너로는 Tomcat v7.0을 이용하여 Database 구축 기반을 마련하였다. <Table 2.1>에 본 연구를 수행하기 위한 시스템 개발환경을 요약하였다. 또한 웹 인터페이스 BLAST 서버 구축을 위하여 NCBI의 WebBLAST 패키지 `wwwblast (ncbi-blast-2.2.26, for linux)`를 설치하였다. BLAST에 연동되는 데이터베이스는 앞에서 유전자 서열을 재가공하여 fasta 형식으로 만든 각각의 유전자 서열을 기반으로 하였다. 각각의 유전자 서열에 포함되어 있는 데이터의 개수는 <Table 2.2>에 나타내었다. 사용자는 비교 하고 싶은 유전자 데이터베이스를 선택하고 쿼리 서열을 입력하게 되면 bit-score 와 e-value를 계산하여 상동성이 높은 상위 100개의 종에 대해 Accession 넘버와 그 종이 포함된 Fungi의 분류체계 결과를 볼 수 있도록 하였다. 그리고 쿼리 서열과 데이터베이스에 존재하는 6개의 Genetic Marker와의 서열을 비교하기 위해 blast 2 sequences program을 install 하여 pair wise alignment가 가능하도록 하였다.

Table 2.1 System development environment. Using the Apache Web server on the Linux operating system, the foundation for building a database, database management system, Mysql, programming language are based on JSP, HTML, and JAVA respectively.

Category	System development environment
System	HPC cluster system
CPU	Master node(24 core), 10 compute node(1 for 8 core)
Memory	Master node(16GB), 10 compute node(1 for 8 GB)
Operating system	Linux
Web server	Apache
DBMS	MySQL
Programming language	JSP, HTML, JAVA

Table 2.2 Amount of data constructed on BLAST database. BLAST database constructed based on genetic marker that usually used for classification of fungi.

Genetic marker for BLAST database	Number of data
rRNA gene	71,853
β -tubulin gene	26,295
Elongation factor 1- α gene	23,116
RNA polymerase subunit II gene	8,810
Atpase subunit six gene	1,478
Calmodulin gene	55

3. 데이터베이스 구현

데이터베이스의 구현은 이를 이용하는 연구자가 데이터를 쉽게 얻을 수 있을 뿐만 아니라 이차 연구에 이용될 수 있는 웹 공간을 제공하도록 설계해야 한다. 이를 위해 데이터베이스의 구현과정은 매우 중요하다. 본 연구에서는 진균류의 유전자 서열을 쉽게 얻을 수 있는 웹 공간 제공과 쿼리 유전자 서열로 진균류의 동정을 할 수 있는 데이터베이스 구현에 목적이 있다. 데이터베이스에는 6개의 마커 유전자 서열을 검색할 수 있다. 각각의 마커 유전자를 선택하면 테이블이 출력되는데 테이블에는 Accession 넘버와 진균류 종의 이름, 계통분류학적 정보가 포함되어 있다. 여기에서 원하는 유전자 서열을 확인하려면 Accession 넘버의 입력을 통해 fasta 형식으로 얻을 수 있다. 뿐만 아니라 Multigene을 생성하기 위해 단일 유전자 마커의 조합이 가능하도록 하여 각 유전자 마커를 선택하면 분류에 이용 가능한 종 정보가 출력되고 선택한 유전자 서열을 모두가 가지고 있는 종들의 서열을 얻을 수 있도록 하였다.

<Table 2.3>에서는 본 연구에서 이용된 데이터들이 데이터베이스 내의 테이블에 어떠한 형태로 저장되었는지 각 테이블 명과 테이블에 따른 필드 명, 데이터 타입을 나타내었다. 테이블에 저장되어 있는 데이터는 모두 문자형이며 종 이름, Accession 넘버, 분류체계 정보는 Varchar 데이터 타입으로 저장하였으며 길이가 매우 긴 유전자 서열 정보는 더욱 많은 문자형 데이터를 저장할 수 있는 text 데이터 타입으로 설정하였다. 또한 Multigene 생성을 위해 각 유전자 별 테이블에서 동일한 종 정보 출력을 위하여 Primary key를 Accession 넘버로 설정하였다.

<Figure 2.3>는 구축한 데이터베이스의 데이터 흐름을 모식도로 표현하였다. 본 데이터베이스의 구현은 크게 세가지로 나뉘어지게 되는데 첫 번

제는 진균류의 계통 분류학적 정보 제공이다. 진균류는 크게 7개의 Phylum으로 나뉘어 지게 되는데 각각의 종마다 형태학적 특징에 유전학적 특징을 추가하여 종의 분류체계를 가지게 된다. 하지만 이러한 분류체계는 현재까지도 유전자 서열에 기반한 분류체계를 더욱 신뢰도 있는 정보로 이용되며 차츰 변화하고 있어 정확한 분류체계 마련은 과거 형태학적 특징으로 분류되었던 진균류의 계통분류 체계를 재 정립 할 수 있을 것이다. 두 번째로는 BLAST를 이용한 진균류의 동정 및 서열 기반 상동성 비교에 있다. BLAST로 연동되는 웹 기반에서 각 유전자 별로 데이터베이스가 구축되어 있다. 사용자가 상동성 검색을 시행하고자 하는 데이터베이스를 선택한 후 쿼리 서열을 넣으면 상동성이 높은 종의 분류 정보가 나와 유전자 서열에 기반하여 종의 분류체계를 확인 할 수 있다. 이는 새로운 진균종의 발견 시 짧은 유전자 서열만을 가지고 정확한 종 동정이 가능하도록 하였다. 마지막으로 진균의 새로운 분류체계 정립을 위해 Multigene 데이터를 쉽고 빠르게 조합할 수 있도록 하였다. 구축된 단일 유전자 6개의 유전자 서열 및 종의 계통 분류학적 정보를 포함하여 다양한 유전자 조합을 수행하여 보다 나은 계통 분류 마커를 찾아 낼 수 있다. 완성된 데이터베이스는 <http://labb.snu.ac.kr/fgdb/>에 공개되어 있다.

Table 2.3 Schema of the database. FGDB constructed table among the marker genes that are string type.

a)

Name of table	Name of gene
Rrna	rRNA Gene
Tub	β -tubulin Gene
Tef	Elongation factor 1- α Gene
Atp	Atpase subunit 6 Gene
Rpb	RNA polymerase subunit 2 Gene
Cal	Calmodulin Gene

b)

Name of field each table	Type of data
Accession	Acc, Varchar(100)
Species	Spc, Varchar(100)
Phylum	Phy, Varchar(100)
Subphylum	Sphy, Varchar(100)
Class	Cla, Varchar(100)
Subclass	Scla, Varchar(100)
Order	Ord, Varchar(100)
Sequence	Seq, Text(1000)

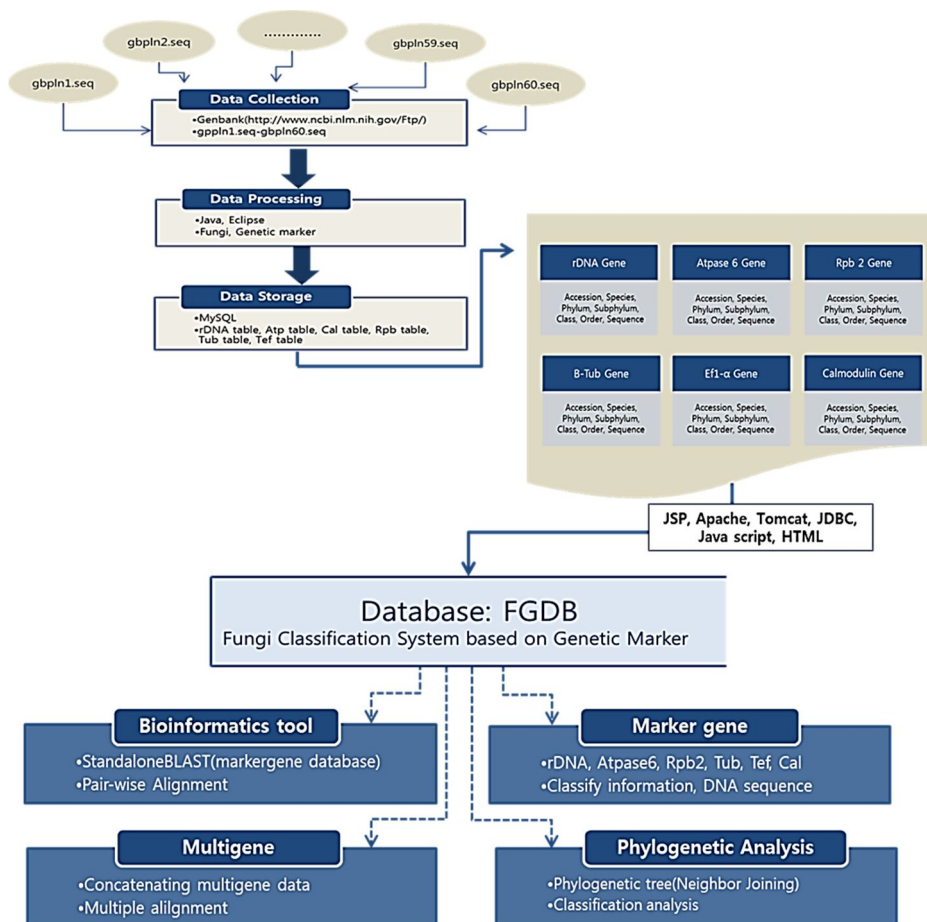


Figure 2.3 Data flow diagram. The database system is able to search the fungal data to provide the identification and the classification information

4. 데이터 분석 항목

본 데이터베이스는 진균류의 유전자 서열 데이터를 통합하여 종의 동정 및 분류체계에 대한 정보를 제공하는데 목적이 있다. 그리하여 데이터베이스로부터 얻은 모든 진균류의 유전자 마커 서열을 바탕으로 상동성 비교 및 검색이 가능하도록 하였다. 데이터로 구축되어 있는 유전자는 진균류의 종 분류에 이용되는 상위 6개의 유전자 마커 rRNA Gene, Calmodulin Gene, Elongation Factor1- α (EF1- α) Gene, RNA polymerase II subunit(RPB2) Gene, ATPase subunit 6 Gene, β -Tubulin Gene를 대상으로 하였다. 유전자 서열 데이터는 얻고자 하는 유전자를 선택하여 종의 계통 분류학적 정보와 함께 얻을 수 있다. 또한 Multigene 조합을 하기 위한 서열 데이터를 선택하면 선택한 유전자들을 모두 가지고 있는 종들의 정보가 출력되며 종별로 유전자 서열을 얻을 수 있으며 유전자 마커들 마다 저장되어 있는 전체 유전자 서열을 얻을 수 있어 단일 유전자 기반의 서열 정렬 후 서열 정렬 결과를 연결하여 Multigene 을 생성할 수 있도록 하였다. <Table 2.4>은 Multigene 조합에 이용한 소프트웨어 프로그램을 나열 하였다.

또한 진균류의 동정 시 서열 비교가 가능하도록 Pair-wise Alignment를 데이터베이스에서 바로 이용할 수 있도록 하였다. 뿐만 아니라 각각의 유전자 서열을 BLAST에서 독립적으로 상동성 검색을 시행할 수 있도록 데이터베이스화 하여 쿼리 서열을 입력하면 상동성이 높은 상위 100개의 종에 대한 유전자 서열과 종의 계통분류학적 정보를 함께 얻을 수 있다. BLAST(Basic Local Alignment Search Tool)는 상동성을 이용한 유전자 검색 방법이다. 이는 NCBI에서 제공하는 데이터베이스를 기반으로 검색을 수행할 수 있으며 독자적으로 구축한 데이터베이스에서 Standalone BLAST 서버 구축을 통해서도 상동성 검색을 할 수 있다. 본 데이터베이스에서도

독자적인 데이터베이스 구축과 함께 BLAST 서버를 구축하여 진균류의 유전자 마커로만 구성된 데이터베이스 안에서 유전자 서열의 상동성 검색을 수행할 수 있도록 하였다. Pair-wise Alignment 는 두 서열 정렬을 통하여 서열의 상동성 분석을 수행할 수 있다. 서열 정렬을 위한 소프트웨어 프로그램은 ClustalX, ClustalW, MUSCLE, MAFFT, SAM, PSAlign 등 의 바이오인포매틱스를 기반으로 한 프로그램이 있으며 본 연구에서는 ClustalX 를 이용하였다(Reblova *et al.*, 2012). 그리고 서열 정렬 결과를 바탕으로 Multigene으로 연결 시켜주는 소프트웨어는 SequenceMatrix1.7.8를 바탕으로 하였으며 계통수를 작성할 수 있는 컴퓨터 프로그램은 PAUP, Phylip, TreeView, MEGA, TreeMap, Phylotools 등 이 있는데 이 중 MEGA를 이용하여 계통수를 작성하였다.

Table 2.4 Analysis systems for making Multigene. Analysis systems used in this study are: the MEGA program for drawing phylogenetic tree, SequenceMatrix for concatenating multigene sequence and ClustalX for multiple sequence alignment.

Category	Analysis system
Multiple Alignment	ClustalX1.83
Concatenating Multigene Sequence	SequenceMatrix1.7.8
Phylogenetic Tree	MEGA5.10

5. 계통수 작성과 정확도 측정

본 연구에서는 구축한 검색시스템을 기반으로 다양한 Multigene 조합을 수행하여 계통수(Phylogenetic tree)를 작성하였다. 계통수를 작성하기 위한 다양한 방법들은 크게 두 가지로 나뉘게 된다. 하나는 Distance based method로 종 마다 서열 간의 차이를 distance matrix를 나타내어 그룹 내의 모든 종족 간의 상호 유사도를 계산한다. 각각의 서열 차이에 기반한 이 방법은 서열 간의 돌연변이가 일어날 확률을 계산하지 않았기 때문에 evolution을 설명하지 못하고 taxa 간의 차이를 보여준다. 이러한 distance matrix를 이용하여 rooted tree에 기반한 UPGMA(Unweighted Pair Group Method with Arithmetic mean)와 unrooted tree에서 기반한 NJ(Neighbor-Joining) 기법이 있다. 반면 서열 간의 치환 확률을 고려하여 계산한 Character based method는 evolution을 설명할 수 있으며 ML(Maximum Likelihood)과 MP(Maximum Parsimony) 기법이 있다.

계통수 작성에 이용한 방법은 NJ 방법으로 이 기법은 Unrooted tree인 별 모양 tree에서부터 시작하여 각각 묶인 neighbor의 total branch length가 최소가 되도록 tree를 생성해 나간다. 즉 각 단계마다 가능한 모든 neighbor를 대상으로 branch의 길이를 최소로 하는 쌍을 골라 그룹을 이루게 하여 tree의 topology와 branch length를 알려준다. 또한 종들은 서로 다른 evolution rate를 갖는다는 가정하에 tree를 생성하기 때문에 evolution rate 차이가 큰 종들을 대상으로 계통 분류를 하는데 적합한 방법이다. 또한 UPGMA와는 달리 각각의 종 사이의 거리 차이를 보는 것이 아니라 전체 data set에서의 차이를 보여주기 때문에 group 내에 distance의 범위가 비슷하지 않은 노드가 있더라도 하더라도 정확한 계통수를 복원해 줄 수 있다.

본 연구에서는 각각의 단일 유전자 마커 및 Multigene 조합을 기반으로 계통수를 작성 한 뒤 각 level 마다 얼마나 분류가 잘 되는지 정확도를 측정하기 위하여 아래의 공식을 이용하였다.

$$\text{Accuracy} = \frac{N_g}{T_g - T_u}$$

N_g : The number of grouping species each level

T_g : Total number of grouping species each level

T_u : Total number of ungrouping species each level

또한 계통수 작성 시 bootstrap value를 100으로 지정하였으며 초기 matrix에서 가능한 sample tree를 100번 조합하여 수행하는 것을 의미하며 각 node 마다 나타난 수치 값들 중 하나가 90일 경우 100번 중 90번에서 이와 동일한 결과를 보이는 것을 말한다. 이 값이 높을수록 계통수의 정확도가 높다고 할 수 있다.

III. 연구 결과

1. 마커 유전자 기반 검색 시스템 구축

본 연구에서는 GenBank(<ftp://ftp.ncbi.nih.gov/genbank/>)에 구축되어 있는 모든 진균류의 6가지 유전자 마커 서열을 통합한 데이터베이스 구축을 시도하였다. 뿐만 아니라 독자적인 웹 인터페이스 BLAST 서버를 구축하여 각 유전자 별로 BLAST 데이터베이스를 구축하여 유전자에 따라 종이 알려지지 않은 진균류의 동정이 가능하다. 상동성 검색에 기반한 BLAST는 서열의 일치 정도를 수리적으로 계산하여 서열을 바탕으로 종의 분류정보를 예측할 수 있다. 데이터베이스에는 6개의 유전자 마커를 클릭하면 저장되어 있는 데이터들이 테이블로 출력되게 되는데 Accession 넘버와 함께 종 이름, 계통분류 정보(Phylum, Subphylum, Class, Order)가 포함되어 있다. 여기에서 종의 유전자 서열을 얻고자 할 때에는 Accession Number를 검색하면 유전자 서열을 얻을 수 있다. 저장되어 있는 데이터의 개수는 각 유전자 별로 차이를 보이는데 GenBank에 80%이상 저장되어 있는 rRNA 유전자의 개수가 가장 많았다. 그 이하로는 beta-tubulin 유전자와 translation elongation factor 유전자가 많았다. 또한 분류 체계에 따른 데이터의 개수는 자낭균류가 가장 많이 존재한다. 자낭균류는 진균류에서 약 60%이상 존재한다고 알려져 있다. 그 다음으로 담자균류가 있으며 가장 데이터가 없는 종은 포자충이다. Multigene 을 이용하기 위해서는 각 분류체계마다 단일 유전자 마커들의 유전 정보를 모두 포함해야 되는데 데이터가 일부 종 및 유전자에 편중되어 있다.

<Table 3.1>은 본 검색 시스템을 바탕으로 나온 데이터의 개수를 나타내었다. 파악 한 유전자 데이터의 전체 개수는 총 131,607개이며 rRNA 유

전자의 개수는 71,853개, β -Tubulin 26,295개, EF1- α 23,116개, RPB 8,810개, ATPase 1,478개, Calmodulin 55개이다. Phylum에 따른 데이터의 개수는 *Ascomycota*가 101,571개로 가장 많고 *Basidiomycota* 27,225개, *Glomeromycota* 1,868개, *Zygomycota* 638개, *Neocallimastigomycota* 103개, *Chytridiomycota* 78개, *Blastocladiomycota* 43개, *Microsporidia* 3개이다. 통합한 데이터 중 Fungi 중 명과 유전자 명이 중복되는 서열 rRNA 11,109개, β -Tub 4,171개, EF1- α 3,490개, Rpb 3,293개, ATPase 16개, Calmodulin 7개는 제외하였다.

Multigene 조합에서는 각각의 유전자 마커 세 개씩 조합을 할 수 있도록 하였으며 선택한 유전자 정보를 모두 가지고 있는 종의 분류학 정보와 서열을 얻을 수 있다. 본 데이터베이스를 바탕으로 Multigene 조합을 수행하기 위해서는 첫 번째 단계에서 조합하고자 하는 유전자들을 선택하여 각각의 유전자 별로 서열정보가 fasta 형식으로 되어 있는 파일을 다운받아 사용자의 컴퓨터에 저장한다. 두 번째 단계에서는 데이터베이스에서 연동시켜 놓은 Multiple sequence alignment 프로그램을 이용하여 유전자 별로 서열 정렬을 한 결과를 fasta 형식으로 다시 저장한다. 마지막 세 번째 단계에서는 소프트웨어 SequenceMatrix를 다운받아 유전자 별로 얻은 Alignment 결과를 연결 시켜 하나의 서열 정렬 파일을 얻을 수 있다. 얻은 결과 파일을 바탕으로 Phylogenetic tree를 작성할 수 있다.

Table 3.1 Amount of sequence data. The number of data for each phylum and related genetic marker is shown. Total number of data is 131,607.

phylum \ marker	ATPase	rDNA	Cal	Rpb	Tef	Tub
Ascomycota	459	51162	55	6309	19575	24011
Basidiomycota	996	18946		2301	3165	1817
Chytridiomycota		14		15	16	33
Zygomycota	21			154	328	135
Glomeromycota		1599		7	13	249
Neocallimastigomycota		101		2		
Blastocladiomycota	2	17		7	3	14
Microsporidia						3
Total	1,478	71,853	55	8,810	23,116	26,295

진균류 동정을 위한 마커 유전자 기반 검색시스템은 사용자를 위한 웹 서비스 제공을 위해 <http://labb.snu.ac.kr/fgdb/>에 웹 공간을 마련하고 구축한 검색시스템 명을 FGDB라고 칭하였다. <Figure 3.1>은 구축한 검색 시스템의 메인 페이지로 데이터베이스 구축 목적과 저장되어 있는 데이터 정보, 데이터베이스의 기능에 대해 명시하였다. FGDB의 주요 기능은 첫째, 6개의 단일 유전자 마커 기반의 검색 시스템을 제공한다. 두 번째, 단일 유전자 마커들을 조합하여 Multigene을 생성 할 수 있는 시스템을 제공한다. 세 번째, 각각의 단일 유전자 마커 별로 BLAST 데이터베이스를 구축하여 서열 정렬 시스템을 제공한다. 마지막으로 두 서열을 정렬할 수 있도록 Pair-wise Alignment를 수행할 수 있는 시스템을 제공한다.

Marker Gene 페이지에서는 통합한 6개의 유전자 마커에 대한 서열 및 구축된 데이터의 중 정보, 분류체계 정보를 얻을 수 있다. <Figure 3.2>는 Marker Gene 페이지를 선택하였을 때의 결과와 함께 Rpb 2 유전자를 선택하였을 때 나타나는 테이블 형식과 특정 종의 Accession 넘버를 선택하였을 때의 결과이다. 왼쪽의 선택 메뉴 바 에서 6가지의 유전자 마커 명과 유전학적 기능을 요약하였으며 유전자 마커를 선택하면 FGDB에 저장되어 있는 모든 종들의 이름과 그 종이 속한 분류체계 정보를 확인 할 수 있으며 특정 종의 서열을 얻고자 할 때에는 Accession 넘버를 선택하면 fasta 형식으로 데이터를 저장할 수 있다.



>> Database of Fungi Marker Gene Sequences

- FunGiDB(FGDB) was built to provide sequence data for the classification of Fungi while especially concerning marker genes.
- The main purpose of the database is to identify fungal species with samples from environments using fungal DNA sequences. FGDB offers marker gene sequence information while combining genes into Multigene is also easy. FGDB implemented pair-wise alignment utility using standalone BLAST application additionally to the database

>>Supplied Functionalities

- Uncombined single marker gene sequence database
- Concatenated Multigene data sets
- BLAST pair-wise alignment functions with marker gene queries



Figure 3.1 Front page of the database. FGDB(<http://labb.snu.ac.kr/fgdb/>) is a database for classification of fungi based on genetic markers.

rRNA Gene

ribosomal rRNA
28S, ITS1, 5.8S,
ITS2, 18S

Description:
Predominant material
within the ribosome
so essential for
protein synthesis in
all living organisms.

β-Tub Gene

Efl-α Gene

Rpb2 Gene

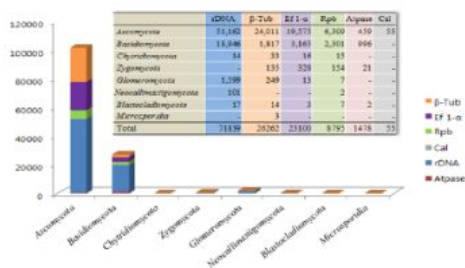
Atpase6 Gene

Cal Gene

Others

6 Marker genes for fungi classification

- Six genetic markers which are usually used in fungi classification**
 - User can download sequence from the result table selecting entries of a species.
 - Downloaded sequence is stored in FASTA format as a separate file
- Information of each marker of all phylogenetic levels was collected from GenBank**
 - User can obtain information from genus to phylum for each species.
 - Data constructed have following number of entries.



	Accession	Species	Phylum	SubPhylum	Class	SubClass	Order
rRNA Gene	HM244776	<i>Abconditella lignicola</i>	Ascomycota	Peizomycotina	Lecanomyces	Ostropomycetidae	Ostropales
β-Tub Gene	HM244778	<i>Abconditella sp.</i>	Ascomycota	Peizomycotina	Lecanomyces	Ostropomycetidae	Ostropales
Efl-α Gene	HM244777	<i>Abconditella iphignorum</i>	Ascomycota	Peizomycotina	Lecanomyces	Ostropomycetidae	Ostropales
Rpb2 Gene	FJ261841	<i>Acanthostichia argentensis</i>	Ascomycota	Peizomycotina	Sordariomycetes	Hypocnemycetidae	Cornophorales
RNA polymerase 2 subunit	AY540532	<i>Acarospora bullata</i>	Ascomycota	Peizomycotina	Lecanomyces	Acarosporomycetidae	Acarosporales
	AY540540	<i>Acarospora canadensis</i>	Ascomycota	Peizomycotina	Lecanomyces	Acarosporomycetidae	Acarosporales
	AY540541	<i>Acarospora cervina</i>	Ascomycota	Peizomycotina	Lecanomyces	Acarosporomycetidae	Acarosporales
	AY540546	<i>Acarospora clauseniana</i>	Ascomycota	Peizomycotina	Lecanomyces	Acarosporomycetidae	Acarosporales
	AY541022	<i>Acarospora complanata</i>	Ascomycota	Peizomycotina	Lecanomyces	Acarosporomycetidae	Acarosporales
	AY540542	<i>Acarospora hirsuta</i>	Ascomycota	Peizomycotina	Lecanomyces	Acarosporomycetidae	Acarosporales
	AY540543	<i>Acarospora laqueata</i>	Ascomycota	Peizomycotina	Lecanomyces	Acarosporomycetidae	Acarosporales
	AY540544	<i>Acarospora macrospora</i>	Ascomycota	Peizomycotina	Lecanomyces	Acarosporomycetidae	Acarosporales
	AY540545	<i>Acarospora schleicheri</i>	Ascomycota	Peizomycotina	Lecanomyces	Acarosporomycetidae	Acarosporales
	AY540546	<i>Acarospora microspora</i>	Ascomycota	Peizomycotina	Lecanomyces	Ostropomycetidae	Ostropales
Atpase6 Gene	FJ218356	<i>Acremonium alternatum</i>	Ascomycota	Peizomycotina	Sordariomycetes	Hypocnemycetidae	Hypocneales
	FJ218354	<i>Acremonium atrogriseum</i>	Ascomycota	Peizomycotina	Sordariomycetes		
	FJ218345	<i>Acremonium breve</i>	Ascomycota	Peizomycotina	Sordariomycetes	Hypocnemycetidae	Hypocneales
Cal Gene	FJ218341	<i>Acremonium</i>	Ascomycota	Peizomycotina	Sordariomycetes	Hypocnemycetidae	Hypocneales

[1] [2] [3] [4] [5] next

생물정보 연구실 서울대학교 생물정보 연구실 TEL : 02)880-2745-2752

147.47.72.71의 HM244776.fasta를 열거나 저장하시겠습니까? 열기(O) 저장(S) 취소(C)

Figure 3.2 Search table of marker gene menu. The markergene menu page shows six genetic markers and offers classification information of species and DNA sequences.

<Figure 3.3>은 Multigene 메뉴를 선택했을 때 모습과 아래에는 ATPase, EF1- α , β -tub를 선택하였을 때 Atpase의 전체 종의 서열을 fasta 형식으로 저장 할 수 있음을 보여준다. 조합하고자 하는 단일 유전자 마커를 선택 하면 종 이름과 유전자 서열을 얻을 수 있는 Accession 넘버를 확인 할 수 있다. 일부 다른 연구에서는 특정 유전자 서열 정보가 없는 경우 공백으로 처리하여 Multigene을 구성하나 본 연구에서는 정확한 유전자들의 조합을 위해 모든 서열 정보가 포함되어 있는 종만을 추출하도록 하였다. 그 결과 3개의 유전자 마커 조합에서는 ATPase, EF1- α , β -tub는 총 107개의 진균종이 포함되어 있고 ATPase, Rpb2, β -tub는 71개, Rpb2, β -tub, EF1- α 는 367개, ATPase, Rpb2, EF1- α 는 220개의 종의 정보가 저장되어 있으며 단계 별로 Multigene 조합을 수행할 수 있도록 하였다. 이를 바탕으로 계통 분류를 시행하면 다양한 마커의 조합을 통해 분류체계의 기준이 되는 Multigene을 찾을 수 있을 것이다.

Step.1

[Atpase6_Efl-α, β-Tub](#)

[Atpase6_Rpb2, β-Tub](#)

[Rpb_Efl-α, β-Tub](#)

[Atpase6_Efl-α, Rpb](#)

Step.2

Step.3

Concatenated Multigene dataset based on FGDB

- Step.1 : Select combination of markers referring information of all phylogenetic level of fungi.**
 - User can download sequence from the result table selecting entries of a species.
 - Whole sequence of a single gene is needed for multiple sequence alignment.
 - Downloaded sequence is stored in FASTA format as a separate file
- Step.2 : Align multiple sequences with various sequence alignment tools.**
 - Data of alignment result should be in FASTA format.
 - CLUSTALW: Thompson, J.D., Higgins, D.G., and Gibson, T.J.1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673-4680.
 - MAFFT: Katoh, K., Misawa, K., Kuma, K., and Miyata, T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*30, 3059-66
 - T-Coffee: Cedric Notredame, Desmond G. Higgins, Jaap Heringa. 2000. T-Coffee: A Novel Method for Fast and Accurate Multiple Sequence Alignment. *J. Mol. Biol.* 302, 205-217
- Step.3 : Concatenating alignment result of data set for phylogenetic analysis.**
 - Sequence matrix could be built from concatenated data set for phylogenetic analysis.
 - The format of the final result could be TNT, Nexus, FASTA and Mega files.
 - SequenceMatrix: Gaurav Vaidya, David J. Lohman, Rudolf Meier. 2011. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics* 27:171-180

Step.1

[Atpase6_Efl-α, β-Tub](#)

[Atpase6_Rpb2, β-Tub](#)

[Rpb_Efl-α, β-Tub](#)

[Atpase6_Efl-α, Rpb](#)

Step.2

Step.3

Spc	Atp Accession	Efl Accession	Tub Accession
Cochliobolus heterostrophus	DQ497603	AB009971	X13439
Nectria haematocystis	AY180204	AF488158	AY489573
Mesocricopea chlamydosporia	DQ223227	AJ012713	EF449508
Cordyceps basiana	AY331948	AJ312228	EU096596
Glomerella citripulvis	AF543772	AJ409291	AY449590
Sterpsia himantoides	AJ518930	AJ518086	AF114450
Coniochaeta arida	AN574581	AN574538	EU335043
Emerizella nidulans	EU119381	AB254358	J01390
Aspergillus niger	F7605512	AB274057	AY232760
Griffithia frondosa	AY385153	AY049143	DQ250688
Trametes versicolor	DQ238603	AY131274	EU335012
Basidiobolus ranarum	AB409158	AY138769	HQ400892
Teratosphaeria caluensis	DQ240128	AY244383	DQ240023
Teratosphaeria molleriana	DQ251504	AY244393	DQ240065
Teratosphaeria multiseptata	AY725582	AY244394	DQ240065
Hymenochaete rufi	AF459004	AY232321	AY489597
Beauveria bassiana	AY331881	AY366059	F9772862

The number of Multigene data: 107 (Multigene Set: Atpase subunit 6, Elongation factor 1-alpha, beta-tub)
Multiple sequence: [Atpase](#) [Efl-alpha](#) [beta-tub](#)

서울대학교 생명정보학 연구실 TEL : 02)880-2745, 2752

147.47.72.71의 Atp.fasta를 열거나 저장하시겠습니까?

열기(O) 저장(S) 취소(C) X

Figure 3.3 Search table of multigene menu. The result of multigene menu page shows three genetic marker combinations and offers species and DNA sequence for the concatenating multigene.

다음으로 진균류의 쿼리 서열을 입력하여 서열 기반으로 종 동정이 가능하도록 구축한 StandaloneBLAST 시스템 구축 결과이다. 구축한 페이지는 <Figure 3.4>에 나타내었으며 각각의 유전자 마커 서열을 기반으로 데이터베이스를 구축하여 사용자가 쿼리 서열을 비교하고자 하는 유전자 마커 Database를 선택하여 서열 정렬을 통해 종 동정이 가능하다. <Figure 3.5>는 임의의 쿼리 서열 데이터를 fasta 형식으로 입력하고 데이터베이스로 β -tub를 선택하였을 때 BLAST를 실행한 결과이다. 실행 결과 β -tub 데이터베이스 안에서 상동성 검색을 통하여 높은 상동성을 가지는 종의 정보부터 하여 상위 100개의 종 정보와 분류체계, 유전정보를 얻을 수 있다. 쿼리 서열의 BLAST 상동성 실행결과 *B.capnodes* 종의 유전자 서열과 93%의 상동성을 보이며 데이터베이스에서 가장 높은 서열 상동성을 갖는 것을 확인 할 수 있었다. 이를 바탕으로 쿼리 서열의 계통 분류학적 체계를 예측할 수 있을 것이다. 더 자세한 상동성 검색을 수행하기 위해 Pair-wise Alignment 시스템 구축 페이지와 실행결과는 <Figure 3.6>와 <Figure 3.7>에 나타내었다. 특히 병원성 진균은 형태학적, 생태학적, 병리학 적으로 다양한 특징을 나타내기 때문에 분자 생물학적 특징인 유전체 서열을 이용한 분류가 가장 신뢰도 있는 결과를 제공해 줄 것이다. 이에 따라 본 연구에서 구축한 BLAST와 Pair-wise Alignment는 병원성 진균의 계통 분류학적 정보와 함께 새로운 진균종의 발견 시 같은 분류체계에 속하는 진균 종들과의 유전자 서열을 비교함으로써 치료 및 이차 연구를 수행할 수 있는 바이오인포매틱스 툴로 이용될 수 있을 것이다.

HOME	MARKER GENE	MULTIGENE	BLAST	ALIGNMENT
------	-------------	-----------	-------	-----------

NCBI **BLAST** BLAST Entrez ?

Choose program to use and database to search:

Program Database

Enter sequence below in

Set subsequence: From To

- rRNA Gene
- Beta-tub Gene
- EF-1 Alpha Gene
- Rpb2 Gene
- Atpsa6 Gene
- Calmodulin Gene
- CHS1 Gene
- COX1 Gene
- COX2 Gene
- GPDH Gene
- Histone Gene
- Hsp Gene
- IGS1 Gene
- IRI1 Gene
- Nach Gene
- Todolomerase Gene

The query sequence is [filtered](#) for low complexity regions by default.
[Filter](#) ☒ Low complexity ☐ Mask for lookup table only

[Expect](#) [Matrix](#) ☐ Perform ungapped alignment

[Query Genetic Codes \(blastx only\)](#)

[Database Genetic Codes \(tblast\[nx\] only\)](#)

[Frame shift penalty](#) for blastx

[Other advanced options:](#)

☒ [Graphical Overview](#) [Alignment view](#)

[Descriptions](#) [Alignments](#) [Color schema](#)

Comments and suggestions to: < blast-help@ncbi.nlm.nih.gov >

Last modified: April 11, 2013



Figure 3.4 Standalone BLAST web interface. Standalone BLAST is constructed based on six genetic markers. Users select database for the identification of unknown query sequence.

FunGi DB
Marker Gene

HOME MARKER GENE MULTIGENE BLAST ALIGNMENT

NCBI BLAST 2 sequences BLAST Entries ?

This tool produces the alignment of two given sequences using BLAST engine for local alignment.

Program

Sequence 1: from: to:
Enter sequence in FASTA format

or download from file

Sequence 2: from: to:
Enter sequence in FASTA format

or download from file

Parameters used in [BLASTN](#) program only:
Reward for a match: Penalty for a mismatch:

[Matrix](#) Open gap and extension gap penalties
Gap x_dropoff [Expect value](#) Word size [Filter](#) ☒

Comments and suggestions to: blast-help@ncbi.nlm.nih.gov

Figure 3.6 Pair-wise alignment web interface. Pair-wise alignment can be used for detailed results compared to BLAST.

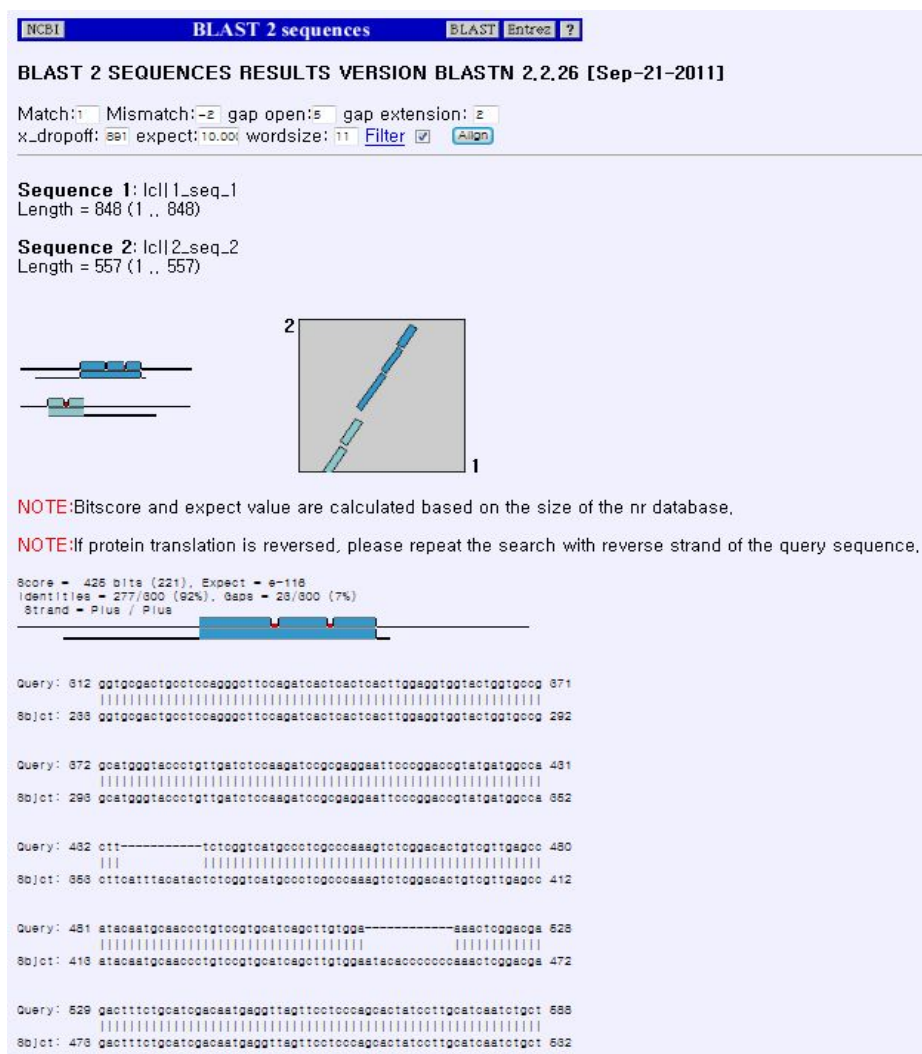


Figure 3.7 Result of pair-wise alignment. The result of pair-wise alignment is shown (with tub gene of *B. capnodes*).

2. 데이터베이스 기반 Multigene 마커 탐색

진균류의 형태학적 특징에 따른 분류의 한계를 보완하기 위해 유전자 서열을 기반으로 한 종 분류는 필수적이다. 과거 rRNA 유전자를 기반으로 생물 종을 크게 원핵생물, 진핵생물, 박테리아 등으로 분류를 할 수 있었던 것을 바탕으로 현재에는 많은 종에서 유전자 서열을 기반으로 한 종 분류체계가 정립되고 있다. 진균류 또한 유전자 서열을 이용하여 분류하기 어려운 종의 계통정보를 알아낼 수 있다. 주로 이용되는 유전자 마커는 rRNA 유전자인데 이 부위 안에 LSU, SSU, ITS 부위 등에 따라 분류 될 수 있는 종이 다르다. 그리하여 다양한 유전자 마커들이 진균류를 분류하는데 있어 발견되었고 이들을 이용한 연구들이 많이 있다. 하지만 현재까지 밝혀진 유전자 마커로는 진균류의 정확한 분류에 한계를 가진다. 또한 다양한 진균류에서는 유전자 마커에 따라 분류가 가능한 종과 불가능한 종으로 나뉘게 된다. 유전자 서열은 진화과정에 따라 차이를 보이고 유전자에서 발현되는 효소 등 단백질 산물들은 진균류의 표현형 및 생활사 등에 영향을 끼치게 된다. 이러한 과학적 이론에 근거하면 진균류를 모두 분류 할 수 있는 통일된 유전자 마커가 있을 것이라고 예측된다. 단일 유전자 마커의 한계로 현재에는 다양한 유전자 마커의 조합을 이용한 Multi-gene Phylogeny에 대한 연구가 활발히 진행되고 있다. Multi-gene Phylogeny는 이용되는 유전자 마커 별로 Multiple Alignment를 수행 한 후 Alignment 결과를 서로 연결해 주는 방법이다.

본 연구에서는 구축한 데이터베이스를 기반으로 하여 전체 진균류를 분류할 수 있는 유전자 마커의 조합을 찾아 진균류의 종 동정 및 분류에 이용할 수 있도록 하였다. 데이터베이스에 저장되어 있는 6개의 유전자 마커들 중 데이터의 개수가 적은 Calmodulin 유전자를 제외한 나머지 5개의 유전자를 대상으로 다양한 조합을 수행하였다. 이용된 진균류 종은 자낭

균류와 담자균류 각 10개의 종에 대해 수행하여 총 20개의 종을 바탕으로 Phylum, Subphylum, Class, Order, Genus를 분류 할 수 있는지 보았다. 이용된 종에 대한 정보는 <Table 3.2>에 정리하였다. 먼저 단일 유전자로 계통분류가 얼마나 잘 되는지 보기 위해 5개의 유전자 마커에 대해 Phylogenetic Tree를 그려 서로 비교해 보았다. 이후 Multigene 기반의 계통 분류학적 연구를 수행하기 위해서 단일 유전자 마커를 Clustalx를 이용하여 Multiple Alignment를 시행한 후 fasta 포맷으로 결과 데이터를 저장하였다. 이후 각 유전자 마커의 Alignment 결과를 SequenceMatrix(v1.7.8)를 이용하여 연결시킨 후 NEXUS 포맷으로 결과 데이터를 생성하였다. SequenceMatrix에 데이터를 삽입한 결과는 <Figure 3.8>에 나타내었다. 계통분류 수행하기 위해 MEGA5.10를 이용하여 Phylogenetic Tree를 Neighbor Joining 방법으로 그렸다. NJ 방법은 서열들의 모든 진화속도가 낮다는 가정하에 유사도에 근거하여 거리를 구하는 것으로 서열상 가장 유사한 것끼리 쌍을 지어 나가는 방법이다. 각각의 Phylogenetic tree 결과는 다음과 같다.

Table 3.2 Species of multigene dataset. Information of 20 species studied in this research is shown (they are in the phylum of Basidiomycota and Ascomycota).

Phylum	Class	Subclass	Order	Genus
<i>Ascomycota</i>	<i>Sordario- mycetes</i>	<i>Hypocreo- mycetidae</i>	<i>Glomerellales</i>	<i>Glomerella</i> <i>Verticillium</i>
			<i>Hypocreales</i>	<i>Balansia</i> <i>Elaphocordyceps</i> <i>Epichole</i> <i>Lecanicillium</i>
			<i>Diaporthales</i>	<i>Diaporthe</i>
		<i>Sodari- omycetidae</i>	<i>Capnodiales</i>	<i>Teratosphaeria</i> <i>Mycosphaerella</i>
			<i>Eurotiales</i>	<i>Penicillium</i>
		<i>Dothideo- mycetes</i>		
<i>Basidiomycota</i>	<i>Agarico- mycetes</i>	<i>Agarico- mycetidae</i>	<i>Polyporales</i>	<i>Grifola</i> <i>Lentinus</i> <i>Ganoderma</i> <i>Trametes</i> <i>Coniophora</i>
				<i>Amanita</i> <i>Flammulina</i>
				<i>Serpula</i>
				<i>Ceratobasidium</i>
				<i>Heterobasidion</i>
			<i>Agaricales</i>	
			<i>Boletales</i>	
			<i>Cantharellales</i>	
			<i>Russulales</i>	

SequenceMatrix 1.7.8							
File Citation Import Export View Sequences							
Delete column	Make outgroup	Delete taxon	Excise entire taxon	Find distances	Display pairwise distances		
Taxon	Total length	No of charsets	atp.fasta	rDNA.fasta	rpb.fasta	tef.fasta	tub.fasta
Amanita	3590 bp	5	737 (5 indels)	752 (49 indels)	990 (78 indels)	575 (37 indels)	536 (45 indels)
Balansia	4127 bp	5	646 (2 indels)	699 (105 indels)	928 (63 indels)	1198 (199 indels)	656 (40 indels)
Ceratobasidium	2802 bp	5	681 (5 indels)	592 (79 indels)	795 (63 indels)	393 (19 indels)	341 (28 indels)
Coniophora	3138 bp	5	730 (11 indels)	737 (56 indels)	626 (101 indels)	584 (40 indels)	461 (40 indels)
Diaporthe	4757 bp	5	515 (2 indels)	1189 (129 indels)	2261 (30 N, 31 indels)	378 (46 indels)	414 (29 indels)
DiaportheEpichloe	2825 bp	1	(No data)	2825 (117 indels)	(No data)	(No data)	(No data)
Elaphocordyceps	5405 bp	5	638 (2 indels)	1817 (164 indels)	1096 (64 indels)	1198 (199 indels)	656 (40 indels)
Epichloe	3486 bp	4	629 (2 indels)	(No data)	1106 (1 N, 64 indels)	1198 (199 indels)	553 (17 indels)
Flammulina	5785 bp	5	746 (5 indels)	853 (54 indels)	2284 (112 indels)	1190 (51 indels)	712
Ganoderma	3968 bp	5	641 (11 indels)	765 (77 indels)	627 (120 indels)	1477 (40 indels)	458 (39 indels)
Glomerella	4057 bp	5	575 (2 indels)	699 (150 indels)	1092 (69 indels)	1198 (199 indels)	493 (38 indels)
Grifola	4120 bp	5	614 (11 indels)	713 (72 indels)	627 (110 indels)	1556 (49 indels)	610 (35 indels)
Heterobasidio	3448 bp	5	593 (12 indels)	701 (84 indels)	627 (101 indels)	1193 (35 indels)	334 (37 indels)
Lecanicillium	6385 bp	5	623 (2 indels)	2825 (526 indels)	1084 (63 indels)	1197 (199 indels)	656 (40 indels)
Lentinus	3392 bp	5	638 (11 indels)	1089 (68 indels)	612 (107 indels)	573 (35 indels)	480 (40 indels)
Mycosphaerella	5140 bp	5	658 (1 N, 2 indels)	699 (166 indels)	2909 (104 indels)	413 (97 indels)	461 (94 indels)
Penicillium	3825 bp	5	191 (6 indels)	1274 (148 indels)	1071 (57 indels)	790 (135 indels)	499 (68 indels)
Serpula	10721 bp	5	647 (11 indels)	8430 (73 indels)	627 (104 indels)	542 (36 indels)	475 (38 indels)
Teratosphaeria	4198 bp	5	658 (2 indels)	1588 (179 indels)	1117 (46 indels)	348 (114 indels)	487 (88 indels)
Trametes	5115 bp	5	647 (11 indels)	754 (94 indels)	596 (107 indels)	1722 (1 N, 4 indels)	1396 (40 indels)
Verticillium	5784 bp	5	683 (2 indels)	2836 (550 indels)	1108 (70 indels)	581 (139 indels)	576 (29 indels)

Figure 3.8 Result of insertion sequence using SequenceMatrix.

<Figure 3.9>는 Rpb 2 유전자 서열을 기반으로 Phylogenetic tree를 작성한 결과이다. 가장 먼저 분지점을 형성하며 나뉘어 져야 할 두 Phylum 그룹이 나뉘어 지지 않은 결과를 보인다. 하지만 자낭균류에서는 하나의 Clade를 형성한 것을 확인 할 수 있다. 또한 자낭균류의 Order level에서 같은 Order에 속한 그룹끼리 Sister group을 형성하였다. <Figure 3.10>은 Efl- α 유전자를 이용하여 Phylogenetic tree를 그린 결과로 Phylum, Class, Order level에서 모두 분류가 제대로 이루어 지지 않았다. <Figure 3.11>은 β -tub 유전자를 이용하여 Phylogenetic tree를 작성하였으며 이 유전자 또한 두 그룹으로 크게 나뉘어져야 할 Phylum의 분류에서조차 정확하게 이루어 지지 않았다. Atpase 6 유전자를 이용하여 Phylogenetic tree를 그린 결과는 <Figure 3.12>에서 확인할 수 있다. 그 결과 Outgroup으로 제일 처음 분지된 *Penicillium*을 제외한 나머지 19개의 종에 대해서는 자낭균류와 담자균류로 정확하게 나뉘어 졌다. 하지만 담자균류의 경우 Class, Order level에서는 분류가 잘 이루어 지지 않았으며 자낭균류에서는 Order level까지 계통분류가 잘 이루어 졌다. 마지막으로 rRNA 유전자를 이용한 Phylogenetic tree는 <Figure 3.13>에 나타내었다. 단일 유전자 마커들 중 분류가 가장 잘 이루어진 rRNA 유전자는 진균류뿐 아니라 박테리아 등에서도 유전자 마커로 주로 이용되고 있다. 다른 단일 유전자 마커들과는 달리 정확하게 두 개의 Phylum으로 초기 분지점을 형성하는 것을 확인할 수 있다. 또한 자낭균류와 담자균류 각각의 그룹 군 안에서도 Class level의 분류가 잘 이루어져 정확한 group을 형성하고 있었으며 자낭균류에서는 Order level까지 동일 그룹 군끼리 cluster를 형성하는 것을 확인할 수 있었다.

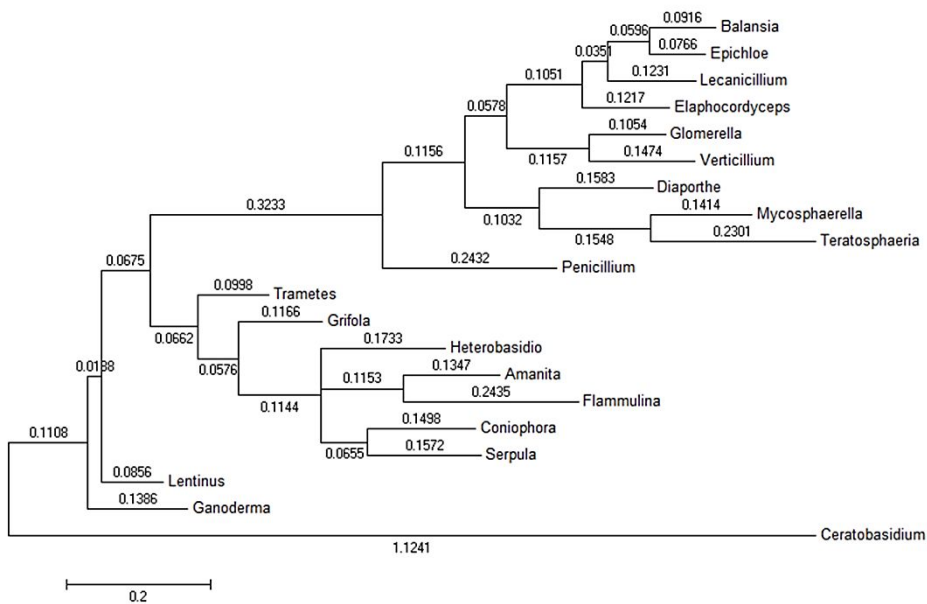


Figure 3.9 Phylogenetic tree(NJ) based on Rpb gene: the result of phylogenetic tree using neighbor joining method based on rpb2 gene sequence.

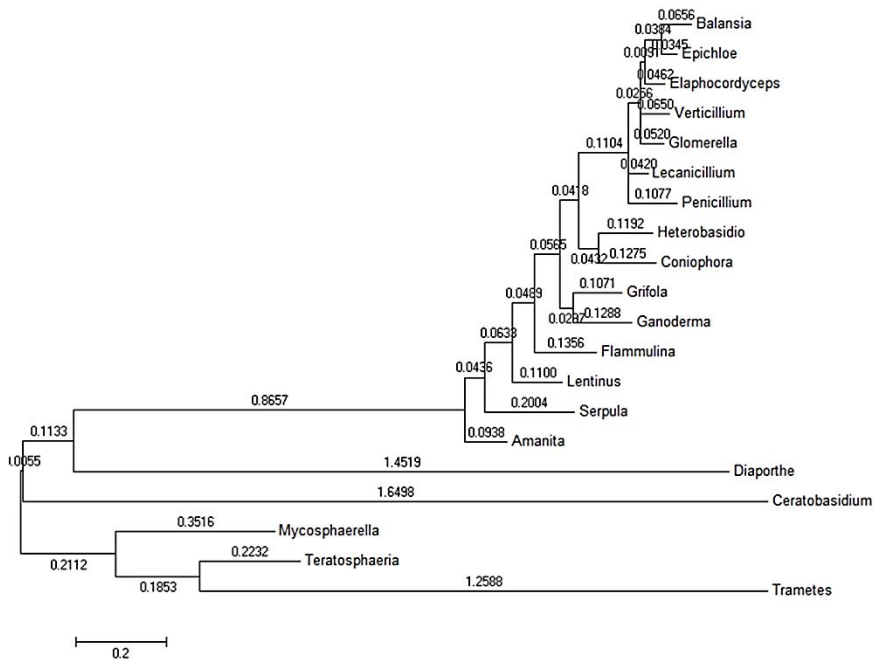


Figure 3.10 Phylogenetic tree(NJ) based on Efl-α gene: the result of phylogenetic tree using neighbor joining method based on ef-1α gene sequence.

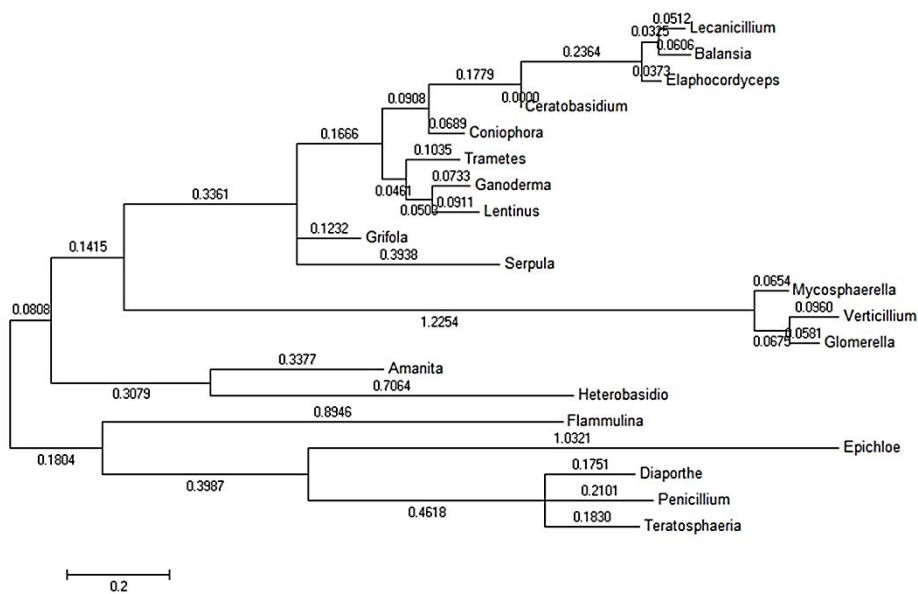


Figure 3.11 Phylogenetic tree(NJ) based on β -Tub gene: the result of phylogenetic tree using neighbor joining method based on β -tub gene sequence.

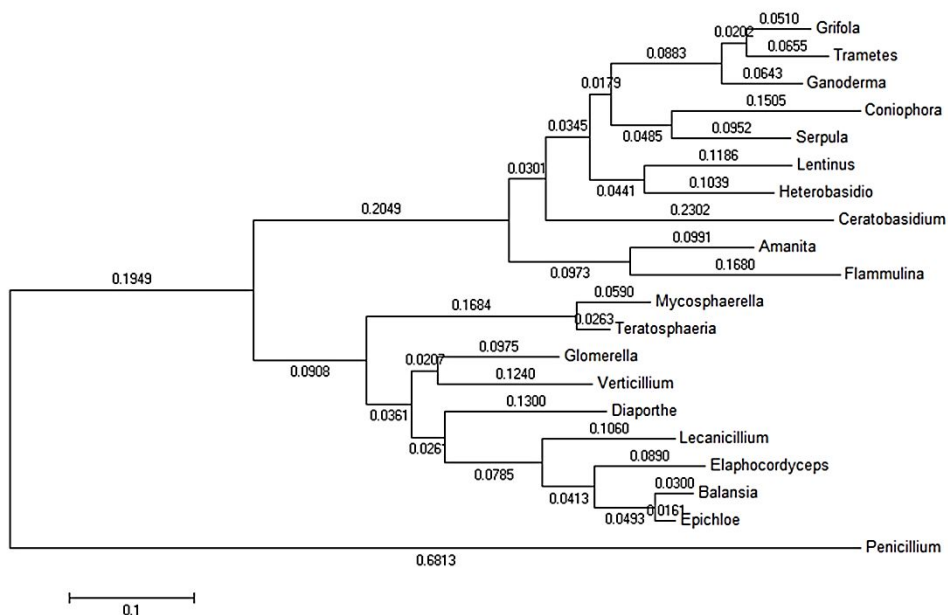


Figure 3.12 Phylogenetic tree(NJ) based on Atpase 6 gene: the result of phylogenetic tree using neighbor joining method based on atpase 6 gene sequence.

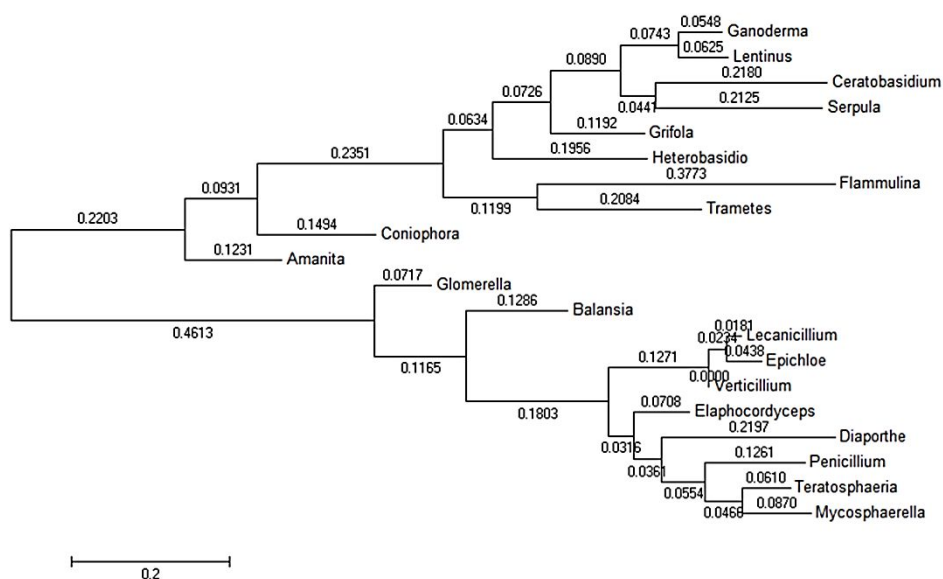


Figure 3.13 Phylogenetic tree(NJ) based on rRNA gene: the result of phylogenetic tree using neighbor joining method based on rRNA gene sequence.

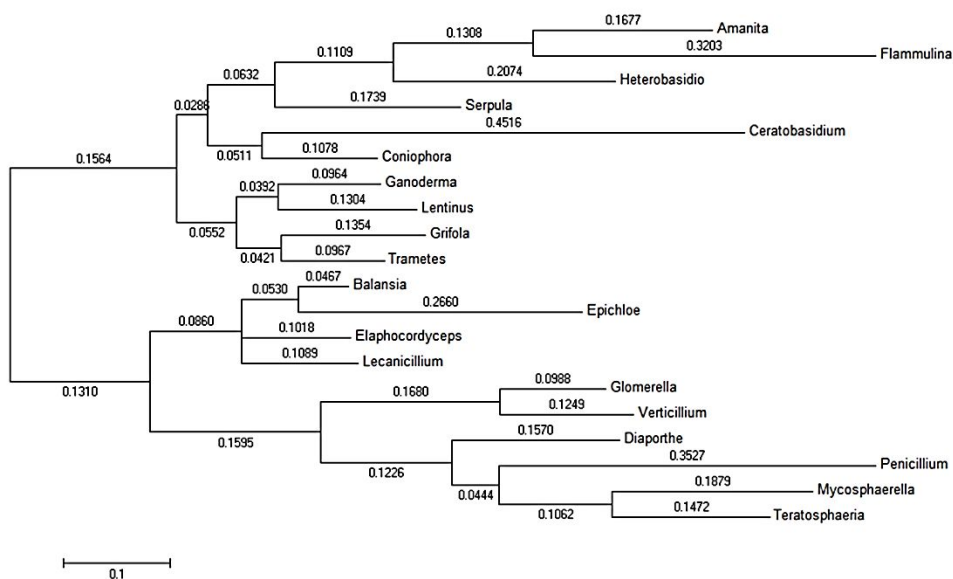


Figure 3.14 Phylogenetic tree(NJ) based on Atpase, Rpb, Tub gene: the result of phylogenetic tree using neighbor joining method based on Atpase subunit 6, RNA polymerase, beta tubulin gene sequence.

단일 유전자 마커 조합을 이용하여 Multigene 생성 후 진균류 분류를 수행한 결과는 다음과 같다. 3개 이상의 유전자 마커의 조합을 수행하였으며 유전자 조합을 하였을 때의 계통분류가 단일 유전자 마커를 이용하였을 때보다 더욱 분류가 잘 되는 것을 확인 할 수 있었다. 또한 유전자 마커를 조합하는데 이용한 유전자의 수가 늘어날수록 계통분류가 잘 되는 것이 아니라 특정 조합에서 전체 종을 분류 할 수 있는 서열 기반이 마련 되는 것을 확인 할 수 있었다. <Figure 3.14>를 보면 atp, rpb, tub 유전자 조합을 이용했을 때 Phylum, Class, Order 분류가 매우 정확하게 이루어 지는 것을 확인 할 수 있다. 자낭균류와 담자균류가 크게 두 개의 branch로 나뉘어 졌으며 각각의 분류 단계에 따라 같은 group으로 묶였다. 이 세 유전자 조합을 이용하였을 때 계통 분류체계의 정확도는 약 92%로 다른 유전자 마커 들 보다 높은 정확도를 보였다. 이외에도 4개의 유전자 마커 조합과 5개의 유전자 마커 조합을 이용하여 Phylogenetic Tree를 작성하였다. 그 결과 <Figure 3.15>는 4개의 유전자 마커 조합(atp+rpb+rRNA+tub)을 이용하면 Phylum의 분류는 잘 되나 일부 Order 이하의 분류가 잘 이루어 지지 않음을 확인 할 수 있다. 또한 <Figure 3.16>에서는 5개의 유전자 마커 조합(atp+rpb+rRNA+tub+tef)를 이용하면 다른 유전자 조합과 마찬가지로 Phylum의 분류는 가능하나 Order 분류가 잘 이루어 지지 않았음을 확인 할 수 있었다. 본 연구에서는 구축한 데이터베이스를 기반으로 Multigene Marker를 이용하여 진균류의 종 분류의 정확도를 비교하여 가장 분류가 잘 이루어지는 유전자 조합을 찾는데 목적을 두었다. 6개의 유전자 마커 중 데이터 수가 가장 작은 Calmodulin 유전자를 제외한 나머지 5개의 유전자 데이터에서 이 5개의 유전자를 모두 가지고 있는 Genus 20종을 골라 계통분류를 수행하였다. 그 결과 단일 유전자 마커로는 가장 큰 분류 단계인 Phylum 인 자낭균류와 담자균류의 분류도 정확하게 이루어 지지 않았다. 뿐만 아니라 유전자에 따라 나타나는 Phylogenetic Tree의 결과도 다 다르게 나타나 통일성을 확인하기 어려웠다. 반면 여러 유전자를

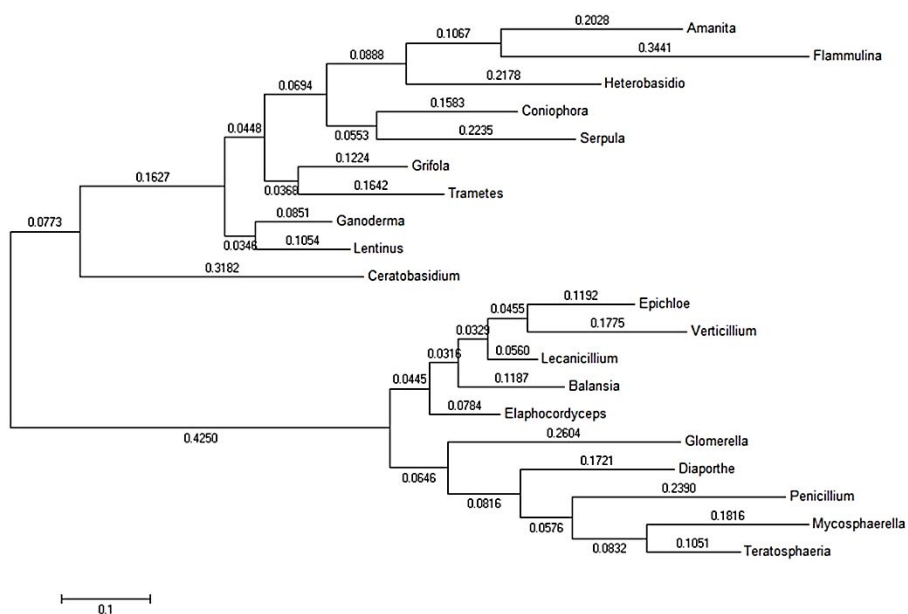


Figure 3.15 Phylogenetic tree(NJ) based on rRNA, Atpase, Rpb, β -Tub gene: the result of phylogenetic tree using neighbor joining method based Atpase subunit 6, RNA polymerase, beta tubulin gene sequence.

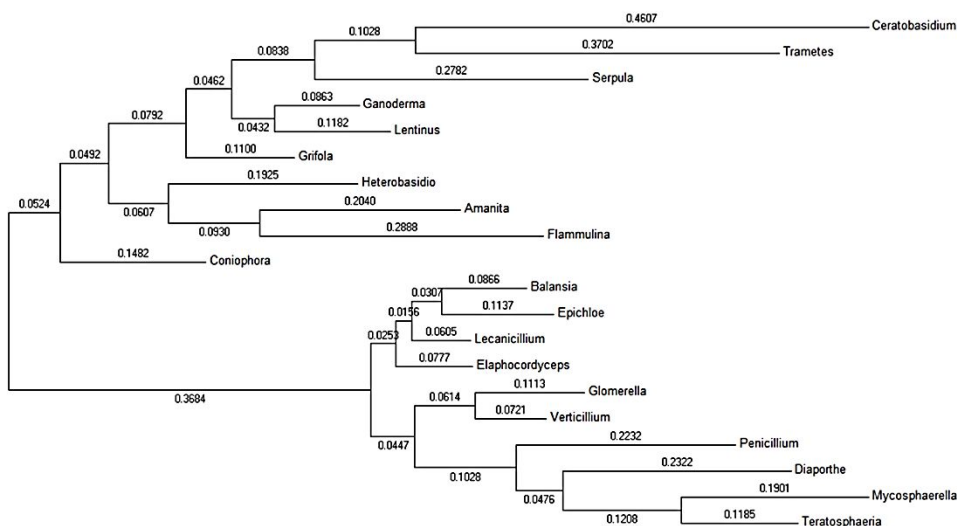


Figure 3.16 Phylogenetic tree(NJ) based on rRNA, Atpase, Rpb, β -Tub, Efl- α gene: the result of phylogenetic tree using neighbor joining method based rRNA, Atpase subunit 6, RNA polymerase, beta tubulin, elongation factor gene sequence.

조합한 Multigene Phylogeny에서는 대부분 자낭균류와 담자균류가 분류되는 것을 확인 할 수 있었다. 이러한 유전자들의 조합은 5가지 유전자를 모두 다 이용하였을 때보다 3개의 유전자 조합인 *atp*, *rpb*, *tub*를 이용하였을 때가 중 분류가 가장 정확하게 이루어 지는 것을 확인 할 수 있었다. 이를 수치적으로 확인하기 위하여 단일 유전자 마커 및 Multigene 조합의 계통수에서 각 분류체계 단계마다의 정확도를 측정하였다. 그 결과는 <Table 3.3>과 같으며 단일 유전자 마커 들 중에서는 rRNA 유전자가 가장 높은 정확도를 가지는 것을 확인 할 수 있었다. 여러 유전자 마커 조합에서는 조합하는 마커의 개수가 증가함에 따라 정확도가 증가하지 않았으며 세 개의 유전자 마커 조합 중 *atp*, *rpb*, *tub*의 조합에서의 정확도가 92%로 단일 유전자 마커 및 이외의 Multigene 조합에서 보다 높게 나타남을 확인 하였다.

Table 3.3 Accuracy of each Phylogenetic tree. Calculating accuracy of each phylogenetic tree using the formula, (The number of grouping species each level)/(Total number of grouping species each level – The number of ungrouping species each level).

level marker	Phylum	Class	Subclass	Order	Family	Genus	Total(%)
Rpb	0.95	0.60	0.60	0.70	0.40	0.60	64%
Efl- α	0.75	0.45	0.40	0.35	0.50	0.75	53%
β -tub	0.55	0.60	0.45	0.40	0.55	0.55	52%
Atpase6	0.95	0.75	0.60	0.60	0.75	0.80	74%
rRNA	1.00	0.70	0.65	0.70	0.80	0.85	78%
Atpase, Rpb, β -tub	1.00	0.90	0.90	0.85	0.85	1.00	92%
Atpase, Rpb, Efl- α	1.00	0.85	0.75	0.80	0.85	0.95	87%
Efl- α Rpb, β -tub	1.00	0.82	0.70	0.85	0.85	0.90	85%
rRNA, Atpase, Rpb, β -tub	1.00	0.85	0.75	0.85	0.80	0.90	86%
rRNA, Efl- α , Atpase, Rpb, β -tub	1.00	0.80	0.75	0.85	0.80	0.90	85%

IV. 고 찰

1. 활용 방안

본 연구의 목적은 진균류의 마커 유전자에 기반하여 서열 검색 및 상동성 비교로 종 동정에 이용될 수 있는 웹 데이터베이스 제공에 있다. 또한 향후 보완을 하여 진균으로 인해 발생하는 질환의 정보 및 관련 진균 종의 정보 제공으로 진균 연구에 기반이 되고자 한다. 따라서 이 검색시스템은 인류의 건강 증진을 목적으로 하는 보건학적 연구와 유전체 분석을 통해 생물의 진화를 연구 및 의학적 연구에 활용 될 수 있다. 현재 진균류를 분류체계 단계별로 정확하게 분류할 수 있는 정확한 단일 유전자 마커 및 Multigene은 아직 밝혀지지 않고 있다. 하지만 다양한 생물들은 진화과정에서 돌연변이 과정을 거치며 서열 변화로 인해 분류체계가 나뉘어 지는데 진균류 또한 유전체 서열을 가지고 있는 생물 군으로 유전자 서열을 기반으로 한 종 분류는 가능 할 것이라고 생각된다. 그리하여 본 데이터베이스는 현재까지 국내외 진균류 연구자들에 의해 밝혀진 유전자 마커 서열을 통합하였다. 이를 바탕으로 종에 대한 계통 분류학적 정보를 모르는 진균종에 대해 상동성 검색을 통해 종의 분류체계를 알 수 있다. 이러한 분류 체계에 대한 정보는 향후 특정 질병을 일으키는 진균류를 동정하여 진화 단계상의 위치를 확인 함으로서 진균증을 치료할 수 있는 의약품 개발에 이용되어 서로 연관이 있는 감염성 진균을 유전학적 메커니즘을 차단하여 감염의 확산을 막을 수 있을 것이다. 앞으로 더욱 많은 진균이 발견될 것이라 예측하고 있는 현 시점에서 현재까지 유전자 서열이 밝혀진 진균을 대상으로 한 데이터베이스의 구축은 필수적이다. 특히 진균증을 일으키는 진균에 대한 유전서열 정보의 통합 및 재정립은 향후 새로운 진균 감염 질환이 나타났을 때 종 동정을 하여 빠르게 치료

할 수 있는 기반이 될 것이다. 또한 진균류에서 페니실린 종과 같은 경우에는 약품의 원료로 이용되기도 하며 푸른 곰팡이의 경우는 발효식품을 생산하는 주요 재료로 이용된다. 이렇게 인간의 생활에 유용한 진균류도 존재하는데 향후 진균류를 이용한 새로운 의약품 개발 및 발효식품이 더욱 다양하게 나올 것이라 예상된다. 이에 따라 진균류의 진화 메커니즘 및 생활사의 기본인 유전학적 정보를 이용한 연구 과정은 필수적이며 본 데이터베이스가 이에 기반이 될 것이라 기대한다. 그리고 FGDB를 바탕으로 새로운 종 동정 및 분류가 가능할 뿐만 아니라 진균류 기초 연구 기반을 제공하며 웹 상으로 제공되는 데이터 베이스는 전 세계 이용자들에게 진균류 분류를 위한 유용한 tool이 될 것이다. 또한 진균증 치료를 위한 신약 치료 및 Drug design에도 이용할 수 있을 것이다.

2. 보건학적 연구에 적용

본 연구에서 구축된 데이터베이스의 보건학적 연구는 주로 진균성 질환을 일으키는 진균증에 대한 것이 주를 이룰 것으로 보인다. 진균증을 완치할 수 있는 치료요법이 부족한 현 시점에서 진균증을 일으키는 진균류들의 유전정보 분석과 발현 메커니즘에 관한 연구는 이에 따른 해결책을 마련할 수 있을 것이다. 그리하여 본 연구에서 구축한 데이터베이스에는 진균증을 포함한 진균류들의 유전정보를 통합하였으며 바이오인포매틱스 기반의 연구가 가능하도록 BLAST를 웹 데이터베이스에서 수행 가능하도록 하였다. 이에 따라 진균증 치료 및 예방의 보건학적 연구는 본 데이터베이스를 이용하면 더욱 효율적인 연구결론을 도출할 수 있을 것이다. 최근 진균증 감염사례가 증가하고 있는데 심한 피부질환을 야기시키거나 면역력이 저하된 환자에게 진균이 감염되어 심하게는 사망에까지 이르게 한다. 아토피 피부염은 전세계적으로 유병률이 증가하고 있는 만성 염증성 피부질환으로 아직까지 완치할 만한 치료법이 존재하고 있지 않다. 이외에도 다른 사람에게 감염을 일으킬 수 있는 진균증의 경우에는 인구집단의 크기가 증가함에 따라 위험에 노출되는 집단도 증가하여 진균증의 발생 빈도 또한 급격하게 늘어난다.

이처럼 보건학적으로 곰팡이 질환에 대한 연구와 진균류의 분류체계의 마련은 인간이 더욱 건강한 삶을 영위하도록 하는 필수적인 연구분야이다. 최근까지의 진균증 관련 연구에서는 진균증이 발생한 환자들을 대상으로 임상학적, 통계학적 연구가 대부분이었다. 하지만 이러한 연구들을 보완하기 위해서는 가장 기초메커니즘인 유전학적 연구가 동반되어야 할 것이다. 진균증은 한번 감염되면 급속도로 다른 부위에 전파되는 경향이 있으며 발생한 이후에는 증상이 완화가 될 뿐 언제든 다시 발병 할 위험이 있는 질환이다. 이에 따라 본 연구에서 구축한 진균류 검색 시스템을 이용하면

연구하고자 하는 진균중에 다른 균류의 유전학적 정보를 얻을 수 있을 것이다. 뿐만 아니라 질병을 일으키는 새롭게 나타나는 진균류들을 대상으로 본 연구에서 발견한 유전자 조합을 기반으로 한 분류체계의 마련은 종 동정을 빠르게 할 수 있어 진균류의 특징 및 감염력 등을 빠르게 파악 할 수 있을 것이다. 또한 향후 추가적으로 질병을 일으키는 메커니즘에 대한 바이오인포매틱스 연구는 현재까지 개발된 항진균제들의 한계를 극복할 수 있는 방향을 마련할 수 있을 것이라 기대한다.

3. 향후 보완 및 추가 계획

구축된 데이터베이스는 현재까지 밝혀진 모든 진균류에서 주로 이용되는 유전자 마커 rRNA Gene, Calmodulin Gene, Elongation Factor1- α (EF1- α) Gene, RNA polymerase II subunit(RPB2) Gene, ATPase subunit 6 Gene, β -Tubulin Gene 의 DNA 서열을 대상으로 하였다. 하지만 이 외에도 Cox Gene, mtSSU Gene, BENA Gene 등 도 진균류의 분류에 이용되는 유전자 마커 중 하나이다. 진균류의 유전자 개수는 종마다 다른데 효모의 유전자 개수의 경우 약 6천여 개가 존재한다고 알려져 있는데 수 천여 개가 존재하는 진균류에서 특정 유전자 서열을 기반으로 한 분류체계의 정립을 위해서는 진균류의 유전자 발견 및 기능에 대한 연구도 필요하다. 이에 따라 향후 본 데이터베이스의 추가계획은 구축한 유전자 마커 이외에도 새로이 이용될 수 있는 유전자 마커를 찾기 위해 더욱 다양한 유전자 서열을 추가 할 것이다. 추가하고자 하는 마커 유전자 대상은 mtrDNA, λ -actin, Rpb1, Cytochrome oxidase 1,2 Histones, GPDH, Nadh, Hsp, Topoisomerase 이다. 이들 유전자들 중에 진균류의 형태학적 분류체계 기반이 되는 포자의 모양, 색 등에 영향을 끼치는 유전자에 대한 분석은 새로운 계통분류학 체계의 정립을 마련할 수 있을 것이다. 또한 더욱 다양한 유전자 마커의 조합을 수행하여 계통분류 체계 마련에 있어 초기 연구 결과보다 더욱 높은 정확도를 보이는 조합을 찾도록 할 것이다. 그리고 추가적으로 퀴리 서열과 함께 저장되어 있는 데이터에서 상동성이 높은 종들과의 Phylogenetic Tree의 결과를 제공하여 퀴리 서열의 진화적 정보를 더욱 체계적으로 알 수 있도록 할 계획이다. 계통수는 생물의 유전 서열을 바탕으로 진화 결과 및 다른 분류군 사이 연관성을 보기 쉽게 다이어그램으로 나타낸다. 이를 위한 소프트웨어들이 많이 개발되어 있으며 이를 본 데이터베이스와 연동하여 퀴리 서열의 종 동정 결과와 함께 나타나도록 하여 종의 진화

관계를 밝히는데 더욱 유용하게 이용될 수 있을 것이다.

본 연구의 목적은 진균류의 유전자를 통합하여 보건학적 이용에 있다. 그리하여 차후 진균증을 일으키는 진균류의 유전정보만을 따로 검색 할 수 있는 시스템을 추가하고 진균증에 따른 증상 및 치료법 등에 대한 정보를 함께 구축할 것이다. 이를 바탕으로 진균증을 일으키는 종의 분류체계 별로 항진균제의 Target이 되는 부위를 찾아내고 이를 바이오인포매틱스 분석 툴을 이용하여 새로운 항진균제 개발에 이용되도록 할 예정이다. 이로써 본 데이터베이스는 진균류의 동정 및 계통 분류학적 정보를 얻을 수 있을 뿐만 아니라 진균증에 대한 정보를 통해 보건학적 연구 이용에 바탕이 될 것이다.

4. 기대성과

진균류에 관련된 데이터베이스 구축은 국외에서 더욱 활발히 진행되고 있다. 이러한 데이터베이스들은 늘어나고 있는 진균류의 유전학적 정보 및 형태학적 특징에 따라 데이터를 통합하여 응용연구에 이용되고 있다. 본 데이터베이스 또한 진균류를 체계적으로 분류할 수 있는 유전자 서열 기반의 데이터베이스로 새로운 진균류 발견 시 데이터베이스를 통해 쉽게 동정할 수 있을 것이다. 또한 진균류 연구에 있어 세계적인 네트워크 연결을 통하여 전 세계적으로 많은 연구진들이 커뮤니케이션 할 수 있는 웹 공간을 제공해 주고 연구자 간 정보 공유 및 연구 소통기반을 형성함으로써 국내의 바이오인포매틱스 분야의 위상을 높일 수 있을 것이라 생각된다. 뿐만 아니라 급증하고 있는 진균증에 대해 진단 및 치료 기법 연구의 토대가 되어 진균류에 대한 유전 서열을 바탕으로 다양한 분석 및 활용이 가능 할 것이다. 또한 새로운 마커 유전자의 발견은 수많은 진균류를 단시간 내에 동정할 수 있을 것이다.

본 연구의 기대 성과는 크게 세가지로 나누어 볼 수 있는데 먼저 학술적 측면에서의 기대성과로는 진균류 연구에 있어 형태학적, 분자 생물학적 연구 방법을 넘어 바이오인포매틱스 기술을 이용하여 새로운 연구 결과를 도출 할 수 있을 것이다. 특히 분자생물학적 연구로부터 얻은 데이터를 기반으로 다차원 연구를 할 수 있을 것이라 기대한다. 또한 산재되어 있는 진균류의 유전자 서열 정보를 통합하여 체계적인 분석을 가능하게 하여 많은 진균학 연구자간의 빠른 정보 공유를 통해 시간 및 노력을 감소 시킬 수 있을 것이다. 다음으로 기술적 측면에서의 기대효과는 본 연구로부터 완성된 데이터베이스를 지속적으로 업데이트 및 관리를 하여 향후 새로운 데이터베이스를 구축할 필요가 없도록 기술적 낭비를 줄이고 바이오인포매틱스 기법을 제공함으로써 전산학적 기반의 연구가 가능해

질 것이다. 또한 새로운 유전자 마커의 개발은 진균류뿐 만 아니라 다른 종에도 적용할 수 있는 가능성을 제시해 준다. 마지막으로 경제, 산업적 측면에서는 데이터베이스의 사용자 수가 증가 함에 따라 국내 과학 기술에 대한 신뢰도와 가치가 상승할 것으로 보이며 진균류를 이용한 다양한 산물 제조 및 의약품 연구의 기반이 될 것이라 기대한다. 따라서 본 연구에서 구축한 마커유전자 기반 진균류 검색 시스템은 진균류의 유전정보를 통합하고 진균류의 계통분류 체계 마련뿐 만 아니라 보건학적으로 문제가 되고 있는 감염성 진균류를 치료할 수 있는 의약품 연구에까지 다양하게 이용될 수 있어 그 기대가치가 매우 크다.

V. 결 론

분자생물학의 발달로 인해 진균류의 유전자 서열의 축적은 데이터베이스의 필요성을 증가시켰다. 뿐만 아니라 환경적 요인으로 인한 진균증의 감염 및 면역치료로 인한 진균의 이차 감염 사례가 급증하고 있다. 피부 표면에서 발병되는 표재성 진균증의 증상은 대부분 가벼우나 인체 내부 또는 전신성 진균증인 경우는 매우 심각한 사례가 많다. 이 때문에 질병을 궁극적으로 인류의 건강을 증진시키는데 목적이 있는 보건학은 진균류에 대한 연구는 중요하다. 본 연구에서는 이러한 연관성을 토대로 진균류의 유전자 서열 데이터를 수집하여 JAVA 프로그래밍 언어를 이용하여 데이터를 재 가공 하여 BLAST 데이터베이스로 이용될 수 있을 뿐만 아니라 유전자 서열과 계통 분류학적 정보만을 얻을 수 있도록 하였다. 수집한 수 만개의 유전자 서열을 관계 형 데이터관리 시스템인 MySQL을 이용하여 저장하였으며 자바 프로그램 안에서 SQL 을 실행하기 위해 데이터베이스를 연결해 주는 응용프로그램 인터페이스인 JDBC를 설치하였다. 마지막으로 데이터들을 웹 공간으로 보여주기 위해 자바를 서버환경에서 사용할 수 있도록 하는 스크립트 방식의 언어인 JSP와 자바 스크립트, 웹 문서를 만들기 위해 기본적인 언어인 HTML을 이용하였다. 바이오인포매틱스 분석 도구로는 웹 인터페이스 BLAST 서버를 구축하여 독자적으로 만든 BLAST 데이터베이스를 이용하여 마커 유전자를 선택하여 특정 데이터베이스 기반의 서열 상동성 검색이 가능하도록 하였으며 이를 웹 서비스가 가능한 진균류 동정을 위한 마커 유전자 기반 검색시스템을 최종적으로 구축하였다.

검색 시스템 구축에 앞서 진균류의 마커 유전자의 종류 및 기능을 파악하기 위해 국, 내외 문헌고찰을 수행하였다. 문헌고찰 결과 진균류의 계통 분류에 이용되는 유전자 마커는 약 20여개 정도이며 주로 이용되는 것은

rRNA 유전자의 LSU, SSU, ITS, 5.8S 부위인 것을 확인할 수 있었다. rRNA 유전자는 단백질을 합성하는 리보솜을 생성하는 매우 중요한 역할을 수행한다. 또한 이 유전자는 모든 세포에 존재하기 때문에 진균류뿐 만 아니라 다양한 생물 종의 계통 분류에 이용되어 왔다. 하지만 rRNA 유전자를 바탕으로 한 진균류의 계통분류에는 특정 그룹 군에만 적용 가능하거나 유전자 서열 길이의 다양성으로 인해 대표적인 유전자 마커로의 이용에 한계를 갖는다. rRNA 이외의 유전자로는 미세 소관을 형성하는 β -tub 유전자와 유전자 전사 과정에 관여하는 Efl- α 와 Rpb 유전자, 에너지 생성을 위해 ATP를 ADP로 분해하는 Atpase 6 유전자 그리고 칼슘의 신호 전달에 관여하는 Calmodulin 유전자가 주로 진균류 분류에 마커로 이용된다. 이에 따라 주로 이용되는 마커유전자를 대상으로 유전자 서열 데이터 및 종의 분류체계 정보를 수집하기 위해 NCBI(National Center for Biotechnology Information)에서 제공하는 유전자 서열 데이터베이스 Genbank에서 FTP 를 통해 gbpln 데이터 전체를 다운로드 하여 Fungi와 각각의 유전자 마커 명을 포함하는 데이터들을 추출하여 Accession 넘버와 종 이름, 분류체계 정보, 서열 데이터만을 파싱하여 각각의 유전자 마커 별로 MySQL에 데이터를 저장하였다. 저장한 데이터들을 바탕으로 유전자 마커에 따라 테이블을 만들고 검색 시스템을 구축하여 웹 서비스가 가능한 데이터베이스를 생성하였다. 그리하여 구축한 데이터베이스 FGDB는 국내, 외 진균류 관련 데이터베이스들과는 달리 유전자 마커 서열들만을 대상으로 하였으며 NCBI에 저장되어 있는 전체 진균류를 대상으로 하였다는 점에서 차별성을 둔다.

다음으로 FGDB에 저장되어 있는 유전자 마커 서열들을 기반으로 다양한 Multigene 조합을 가능하게 하였다. MySQL에 저장되어 있는 6개의 유전자 마커 테이블은 Primary key를 Accession 넘버로 지정하여 다른 테이블과의 Inner join이 가능하도록 하였다. 즉, 세 개의 테이블에서 종 이름이

일치 하는 데이터들의 Accession 넘버를 출력하게 하고 각각의 Accession 넘버를 선택하였을 때 유전자 서열을 얻을 수 있도록 하였다. 이후 단계 별로 Multigene 생성을 위한 프로그램들을 FGDB의 메뉴에서 연동하게 하여 사용자가 FGDB를 기반으로 Multigene을 조합하여 Phylogenetic tree를 생성할 수 있다. 본 연구에서는 FGDB를 기반으로 단일 유전자 마커들을 조합한 Multigene 마커들을 이용하여 진균류의 분류를 수행하였다. 그 결과 가장 높은 정확도를 보인 유전자 마커들의 조합은 단백질을 코딩하고 있는 유전자 *Atpase*, β -*Tub*, *Rpb*이다. 이 유전자 조합을 기반으로 진균류의 분류 체계를 마련한다면 현재까지 이용된 유전자 마커들의 한계점을 극복할 수 있을 것이다. 하지만 현재 Genbank에 저장되어 있는 진균류의 유전자 서열은 대부분이 rRNA 유전자이기 때문에 *Atpase*, β -*Tub*, *Rpb*의 유전자 데이터들이 빠르게 Sequencing 되어야 할 것이다. 향후 본 연구에서는 추가적으로 현재 구축되어 있는 6가지의 유전자 마커 이외에 진균류 분류에 이용되었던 유전자 마커 데이터를 통합할 계획이다. 이로써 FGDB를 기반으로 진균류 분류체계 정립을 해결할 수 있는 실마리를 찾을 수 있을 것이라 기대한다.

의료기술이 발달 할수록 진균류 관련 질병들은 더욱 심각해 지고 있는 것이 사실이다. 그 이유는 장기이식, 항암 치료 등의 발달로 인해 많은 환자들이 면역력이 약해져 있는 상태에서 진균류에 감염되면 초기 진단이 어려울 뿐만 아니라 심재성 진균증으로 발전되어 장기에 치명적으로 빠르게 확산되어 생명을 위협하기 때문이다. 현재까지 진균증의 치료는 주로 약물치료를 하고 있는데 진균류는 동물 세포와 같은 특징을 갖기 때문에 약물이 인체 세포에까지 악영향을 끼치는 등 부작용을 일으킨다. 이러한 한계를 극복하기 위해서는 새로운 연구 방법이 필요할 것이며 다량의 데이터를 처리하고 분석할 수 있는 바이오인포매틱스 연구가 그 해결책을 제시해 줄 수 있을 것이라 생각한다.

결론적으로, 본 연구는 전산기술을 이용하여 진균류 동정을 위한 마커 유전자 기반 검색시스템을 구축하여 마커 유전자에 따라 종 이름 및 분류 체계 정보와 서열 데이터를 얻을 수 있을 뿐만 아니라 다양한 유전자 조합을 통해 진균류 분류에 이용할 수 있는 **Multigene**을 생성할 수 있도록 하였다. 또한 유전자 마커 각각을 데이터베이스화 하여 BLAST에 적용하여 쿼리 서열을 입력하고 서열 정렬을 하고자 하는 유전자명을 선택하여 종 동정을 가능하게 하였다. 향후 지속적인 데이터 업데이트와 감염을 일으키는 진균류 데이터를 특성화하고 바이오인포매틱스 분석 기법을 보완할 것이다. 이는 진균류 연구의 기반이 되는 데이터베이스로 유용하게 이용될 수 있을 것이라 기대한다.

참고문헌

- Adam L Bazinet, Michael P Cummings.** A comparative evaluation of sequence classification programs. *BMC Bioinformatics*. 2012;13:92-98.
- Anaissie E.** Opportunistic mycoses in the immunocompromised host: experience at a cancer center and review. *Clin Infect Dis*. 1992;14(1):S43-53.
- Andrew J. Roger, Olof Sandblom, W. Ford Doolittle, Herve' Philippe.** An Evaluation of Elongation Factor 1 α as a Phylogenetic Marker for Eukaryotes. *Mol. Biol. Evol*. 1999;16(2):218-233.
- Arzanlou M and Khodaei S.** Phenotypic and molecular characterization of *Chaetopyrena penicillata* from Iran with description of a hyphomycete synanamorph *Mycosphere* Doi 10.5943/mycosphere/3/1/9. 2012.
- Baldauf SL, Palmer JD.** Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. *PNAS*. 1993;90(24):11558-11562.
- Blanka Havlickova Viktor A. Czaika, Markus Friedrich.** Epidemiological trends in skin mycoses worldwide. *Mycoses*. 2008;51(4):2-15.
- C. d'Enfert, S. Goyard, S. Rodriguez-Arnaveilhe, L. Frangeu, L. Jones, F. Tekaia, O. Bader, Antje Albrecht, L. Castillo, A. Dominguez, J. F. Ernst, C. Fradin, . Gaillardin, S. Garcia-Sanchez, P. de Groot, B. Hube, F. M. Klis, S. rishnamurthy, D. Kunze, M.-C. Lopez, A. Mavor, N. Martin, I. Moszer, D. One'sime, J. Perez Martin, R. Sentandreu, E. Valentin and A. J. P. Brown.** CandidaDB: a genome database for *Candida albicans* pathogenomics. *Nucleic Acids Research*. 2005;33: D353-D357.
- Chandler F, Watts J.** Mycotic, Actinomycotic, and Algal Infections. In Kissane J (ed) *Anderson's Pathology*. 1996: 391-432.
- Cherry JM, Caroline A, Catherine B, Stephen AC, Selina SD, Erich TH, Yankai J, Gail J, TaiYun R, Mark S, Shuai W, David B.** SGD: *Saccharomyces* Genome Database. *Nucleic Acids Research*. 1998;26(1):73-79.
- Choi JY, Park JS, Kim DH, Jung KY, Kang SC, Lee WH.** Fungal Secretome Database: Integrated platform for annotation of fungal secretomes. *BMC Genomics*. 2010;11:105-115.

Clinical and Laboratory Standard Institute. Interpretative criteria for identification of bacteria and fungi by DNA target sequencing. approved guideline. Pennsylvania, Clinical and Laboratory Standard Institute, 2008; MM18-A:1-71.

Cruz Ch R, Ponce E E, Calderón R L, Delgado V N, Vieille O P, Piontelli L E. Superficial mycoses in the city of Valparaiso, Chile: Period 2007-2009. *Rev Chilena Infectol.* 2011;28(5):400-409.

David W. Fraser, Joel I. Ward, Libero Ajello, Brian D. Plikaytis, MS. Aspergillosis and Other Systemic Mycoses The Growing Problem. *JAMA.* 1979;242(15):1631-1635.

Elewski BE, Hazen PG. The superficial mycoses and the dermatophytes. *Journal of the American Academy of Dermatology.* 1989;21(4):655-673.

Jim Deacon. *Fungal Biology.* 4th ed. World Science. 2006.

Flavio Queiroz-Telles, MD, Michael R. McGinnis, MD, Ira Salkin, MD, John R. Graybill, MD. Subcutaneous mycoses *Infect Dis Clin N Am.* 2003;17:59–85.

Frank MD, George N. New records for powdery mildews and *Taphrina* species in Idaho and Washington. *Pacific Northwest Fungi.* 2007;2(8):1-5.

Fryar SC, Booth W, Davies J, Hodgkiss IJ, Hyde KD. Distribution of fungi on wood in the Tutong River,. Brunei. *Fungal Divers.* 2004;17:17–38.

Gazis R, Rehner S, Chaverri P. Species delimitation in fungal endophyte diversity studies and its implications in ecological and biogeographic inferences. *Molecular Ecology.* In press. 2011;20:3001-3013.

Hainer BL. Dermatophyte Infections. *AMERICAN FAMILY PHYSICIAN.* 2003;67(1):101-108.

Harmesen D, Schwinn M. Weig EB, Heesemann J. Phylogeny and dating of some pathogenic keratinophilic fungi using small subunit ribosomal RNA. *Journal of Medical & Veterinary Mycology.* 1995;33:299-303.

Hawksworth DL. The fungal dimension of biodiversity: magnitude, significance, and conservation. *Mycological Research.* 1991;95:641-655.

Hawksworth DL. The magnitude of fungal diversity: the 1.5 million species estimate revisited. *Mycol. Res.* 2001;105(12):1422-1432.

Hibbett, D.S., *et al.* A higher-level phylogenetic classification of the Fungi. *Mycological Research*. 2007;111:509-547.

Ismael A. Conti Diaz. Epidemiology of sporotrichosis in Latin America *Mycopathologia*. 1989;108:113-116.

James EG, Matthew RH, Li-Jun M. Genomics of the fungal kingdom: Insights into eukaryotic biology. *Genome Res*. 2005;15:1620-1631.

Justyna Karkowska-Kuleta, Maria Rapala-Kozik and Andrzej Kozik Acta. Fungi pathogenic to humans: molecular bases of virulence of *Candida albicans*, *Cryptococcus neoformans* and *Aspergillus fumigatus*. 2009;56(2):211–224.

Kauffman CA. Endemic Mycoses: Blastomycosis, Histoplasmosis, and Sporotrichosis. *Infect Dis Clin N Am*. 2006;20:645-662.

Keith AS. Progress towards DNA barcoding of fungi. *Molecular Ecology Resources*. 2009;9(1):83–89.

Kenneth W, Cullings, Detlev R, Vogler A. 5.8S nuclear ribosomal RNA gene sequence database: applications to ecology and evolution. *Molecular Ecology*. 1998;7:919-923.

Kevin M. The Fungal Genetics Stock Center: From Molds to Molecules. *Advances in Applied Microbiology*. 2009;52:245-262.

Lachke SA, Shawn R. Lockhart, Karla J. Daniels and David R. Soll. Skin Facilitates *Candida albicans* Mating. *American Society for Microbiology*. 2003;71(9):4970-4976.

Landvik S, Eriksson OE, Berbee ML. Neoelecta – a fungal dinosaur? Evidence from beta-tubulin amino acid sequences. *Mycologia*. 2001;93(6):1151–1163.

Leedale GF. How Many are the kingdoms of Organisms?. *Taxon*. 1974;23(2/3):261-270.

Lorenzo HK, Susin SA. Mitochondrial effectors in caspase-independent cell death. *FEBS Letters*. 2004;557:14-20.

Mahreen Ameen, MD. Epidemiology of superficial fungal infections. *Clinics in Dermatology*. 2010;28:197–201.

Maksymiuk AW, Thongprasert S, Hofer R, Luna M, Fainstein V, Bodey GP. Systemic candidiasis in cancer patients. *The American Journal of Medicine*. 1984;77(4D):20-27.

Malcolm D. Richardson. Changing patterns and trends in systemic fungal infection, *Journal of Antimicrobial Chemotherapy*. 2005;56: suppl. S1, i5-i11, doi:10.1093/jac/dki218.

Martina Réblová, Kamila Réblová. RNA secondary structure, an important bioinformatics tool to enhance multiple sequence alignment: a case study (Sordariomycetes, Fungi). *Mycol Progress*. DOI 10.1007/s11557-012-0836-8.

Michael R. McGinnis. Chromoblastomycosis and phaeohyphomycosis" New concepts, diagnosis, and mycology] *JOURNAL Of the American AcaDemy OF DeFMaTOLOGY*. 1983;8(1):54-60.

Nicolas F, Thibaut D, Claude H, Marie-Laure Desprez-Loustau, Cyril D. Finding Single Copy Genes Out of Sequenced Genomes for Multilocus Phylogenetics in Non-Model Fungi. *PLoS ONE*. 2009;6(4):e18803. doi:10.1371/journal.pone.0018803.

Schenck BR. On refractory subcutaneous abscesses caused by a fungus possibly related to sporotrichia. *Bulletin of the Johns Hopkins Hospital* 1998;9:286–90.

Schmitt, A. Crespo, P.K. Divakar, J.D. Fankhauser, E. Herman-Sackett, K. Kalb, M.P. Nelsen, N.A. Nelson, E. Rivas-Plata, A.D. Shimp, T. Widhelm, H.T. Lumbsch. New primers for promising single-copy genes in fungal phylogenetics and systematics. *Persoonia*. 2009;23:35–40.

Seifert K. Progress towards DNA barcoding of fungi. *Molecular Ecology Resources*. 2009;91:83–89.

Selina S. Dwight, Midori A. Harris, Kara Dolinski, Catherine A. Ball, Gail Binkley, Karen R. Christie, Dianna G. Fisk, Laurie Issel-Tarver, Mark Schroeder, Gavin Sherlock, Anand Sethuraman, Shuai Weng, David Botstein and J. Michael Cherry. *Saccharomyces* Genome Database (SGD) provides secondary gene annotation using the Gene Ontology (GO). *Nucl. Acids Res*. 2002;30(1):69-72.

Sharon C-AChen, E Geoffrey Playford, Tania C Sorrell. Antifungal therapy in invasive fungal infections. *Current Opinion in Pharmacology*. 2010;10:1–9.

Silvia Castillo-Argüero, José A. Ramos-Zapata, Sara L. Camargo-Ricalde, Javier Álvarez-Sánchez. Propagules of arbuscular mycorrhizal fungi in a secondary dry forest of Oaxaca, Mexico. Patricia Guadarrama, Rev. Biol. Trop. 2008;56(1):269-277.

Sorel T. Fitz Gibbon and Christopher H. House. Whole genome-based phylogenetic analysis of free-living microorganisms. Nucl. Acids Res. 1999;27(21):4218-4222.

Sylvain M, Gabriela A, François R, Annie G, Tatiana G, Elisabeth F, Manuela LV, Angélique G, Marc-Henri L, Hélène C. FUNYBASE: a FUNgal phylogenomic database. BMC Bioinformatics. 2008;9:456-466.

T.G. Frøslev, P.B. Matheny, D.S. Hibbett. Lower level relationships in the mushroom genus *Cortinarius* (Basidiomycota, Agaricales): A comparison of RPB1, RPB2, and ITS phylogenies. Mol Phylogenet Evol. 2005;37(2):602-618.

Tetsuya K, Tetsuo M, Tadahiko M, Masutaka F. Therapeutic Approaches to Subcutaneous Mycoses. 2003;4(8):537-543.

Yuuhiiko T, Makoto M. Watanabe, Junta S. Are *Microsporidia* really related to *Fungi*?: a reappraisal based on additional gene sequences from basal fungi. Mycological Research. 2002;106(12):1380-1391.

Abstract

Fungi Classification Based on Genetic Markers

Name : Jang, Jin-Hwa

Department and Major : Bioinformatics,

The Graduated School of Public Health,

Seoul National University

Fungi are parasitic on host cells and perform the role of decomposers of the ecosystem in the environment in which various organisms coexist. They are practically used for food ingredient and are also unwanted disease agents in mycoses. In recent years, outbreaks of infectious fungi are increasing, due to the changes in living environment. Also the frequency of occurrence of fungal diseases is increasing, which will become major public health problem especially in the case of strong contagion. Thus, the studies on the genetics and the life cycle of fungi became essential not only in biology but also in public health studies. Population of fungal species is estimated about 1.5 million while only about 100,000 species have been DNA sequenced so far.

However, recent development of molecular biology and NGS (Next Generation Sequencing) technology result in rapid accumulation of sequence information for fungal species. Constructing and maintaining of databases are undoubtedly essential because of newly discovered fungi needed to be systematically characterized. Characterization and classification are prerequisite to further studies of every fungal research. Identification and utilization of reliable genetic markers are also therefore important for the best characterization and the classification of the species. In this study, an information search system based on genetic markers has been designed and constructed for the identification of fungal species. The search system, named as FGDB, is linked to the genetic marker database for the identification of the species. The database has integrated meta-information about classification and sequences of the genetic markers that are commonly in use. FGDB is accessible on the internet through the web site <http://lcbb.snu.ac.kr/fgdb/>. By using the system, users can get information and sequence classification scheme based on genetic markers of fungi. Also when users supply query sequences to the BLAST search module, top 100 highly homologous results will be returned. Using the system, a research was conducted to find the best combination of multigene markers for effective fungal classification. As a result, it was shown that combination of Atpase6, Rpb2 and β -Tub genes is the most reliable for the classification. This study is a firm basis for the future research and it is hoped that research method reported will be further developed for practical use in the field of public health.

Keywords : fungi, mycoses, bioinformatics, database, phylogenetic analysis, genetic marker, public health.

Student number : 2011-23863

감사의 글

2011년 하얗게 눈이 내렸던 겨울, 연구실에 처음 발을 디뎠던 때가 며칠 전 일 인 것처럼 느껴지는데 벌써 졸업논문을 마무리하고 감사의 글을 쓰는 이 순간이 믿겨지지 않습니다. 저에게 있어 졸업논문은 석사 과정 동안 많은 것을 배우며 얻은 학문적 지식뿐만 아니라 저에게 더욱 높은 꿈을 꿀 수 있게 만들어 주었습니다. 본 졸업논문을 완성하는데 많은 지도를 해주시고 연구뿐 만 아니라 참된 과학인으로서 갖추어야 할 많은 지식과 우리나라를 이끌어 가야 할 학자로서의 리더십, 통찰력 등 저에게 많은 가르침을 주신 손현석 교수님께 깊은 감사의 말씀을 드립니다. 교수님의 가르침을 받들어 우리나라를 위한 훌륭한 연구자와 리더로서 큰 역할을 하는 자랑스런 제자가 되도록 노력하겠습니다. 그리고 뒤에서 언제나 저의 든든한 지원군이 되어 주셨던 아버지, 어머니께도 감사의 말씀을 드립니다. 앞으로 더욱 부모님의 자랑스런 딸이 되도록 하겠습니다. 비록 학위과정은 긴 시간은 아니었지만 그 사이 저는 많은 성장을 하였습니다. 이러한 성장 과정은 저 혼자 이룬 것이 아니라 같이 연구실 생활을 하였던 좋은 선배, 후배님들과 함께 하였기에 더욱 이 순간이 기쁘고 뿌듯합니다. 멀리에서도 언제나 후배들을 위해 많은 도움을 주신 안인성 박사님과 특히 저에게 있어 학문적 소양을 쌓는데 큰 힘이 되어 주시고 연구실의 많은 고민을 함께 나누어 주신 배세은 교수님께 감사 드립니다. 또한 멋지게 사회생활을 하고 계신 김하연 선배님, 든든한 정성훈 박사님, 연구실의 큰 언니 김미란 선배님, 본받을 점이 많은 이지혜 선배님, 힘들 때 함께 하며 웃음을 주었던 미경씨랑 명지씨 그리고 이제 막 연구실 생활을 시작한 서영씨까지 저와 함께 하며 기쁨과 슬픔을 함께 나누며 저에게 그 무엇보다 값지고 소중한 것들을 얻을 수 있게 해 주신 교수님과 연구실 식구들에게 이 글을 빌어 감사하다고 전하고 싶습니다. 마지막으로 그 동안의 노력의 결실을 맺고 새로운 시작을 앞두고 있는 저에게 더욱 큰 꿈을 품고 한걸음 더 나아가기 위한 용기와 격려의 박수를 보냅니다.